Evaluation Of SARS-CoV-2 Entry Factors In Chronic Lymphocytic Leukemia Via Bioinformatics Analysis

Abdulmohsen M Alruwetei

1Medical Laboratories Department, College of Applied Medical Sciences, Qassim University, Qassim, Saudi Arabia
roietaie@qu.edu.sa

Abstract: Coronavirus disease of 2019 (COVID-19), which caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is recognized as a life-threatening disease accounting for a considerable large morbidity and mortality in Chronic Lymphocytic Leukemia (CLL) patients. The interplay between the virus and CLL disease at molecular level and whether the biology of CLL disease can contribute to the severity of infection are uncertain. Herein, I investigated the gene expression of 15 SARS-CoV-2 and coronavirus-related receptors and factors in CLL disease based on meta-analysis of microarray datasets. The analysis revealed significant increase in the expression of 4 entry-promoting factors (BSG, CLEC4G, CLEC4M and DPP4), whereas the well-known entry receptor (i.e. ACE-2) and protease (i.e. TMPRSS2) for SARS-CoV-2 were not expressed at significant high level in CLL disease. In regard to the entry-restricting factors, CLL disease demonstrated a significant decrease in IFITM2 and LY6E genes. These results reiterated the findings from other studies which indicated a higher gene expression of the virus-entry receptors as well as lower gene expression of the virus-entry restrictors in COVID-19 comorbidity diseases compared to healthy controls. These findings suggest possible roles of the host genetic factors related to SARS-CoV-2 entry on the comorbidity between CLL and COVID-19 disease.

Key words: Chronic Lymphocytic Leukemia, CLL, COVID-19; SARS-CoV-2 entry factors

1. INTRODUCTION:
The Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), a novel beta lineage-CoV, is the causative microorganism of a worldwide pandemic and highly infectious disease, named as Coronavirus-2019 (COVID-19). The disease is characterized by variable susceptibility rate in human and heterogeneous clinical and comorbidity factors which can influence the outcomes among infected people (1, 2). The symptoms of COVID-19 disease range from subclinical or moderate common cold and fever to acute and chronic respiratory symptoms and other serious extrapulmonary manifestations. While most healthy young and children present with mild symptoms and an overall good prognosis, the disease is more severe in adults specifically those with certain comorbidity factors such as cancer (including
hematological malignancies), cardiovascular disease, obesity, type 2 diabetes mellitus and primary and secondary immunodeficiency conditions (3).

There is increasing evidence from the literature that patients with hematological malignancies are at increased risk of serious complications and high mortality rate from COVID-19 disease (4-8). Of particular, Chronic Lymphocytic Leukemia (CLL) patients are expected to have high risk for SARS-CoV-2 infection and subsequent complications and death from COVID-19 disease owing to the predominance of CLL in elderly people and the immune suppression resulted from CLL disease itself and also from the immune-directed therapies for CLL disease (9). Severe complications and death from COVID-19 disease were reported among CLL patients who undergone therapeutic interventions. Patients on watch and wait management who usually do not receive treatment before evidence of disease progression and considered to have a lower risk of severe COVID-19 disease due to relative residual immunity compared to CLL-treated patients (9, 10). The risk factors which could contribute to the high rate of SARS-CoV-2 infection and serious COVID-19 disease in CLL patients could be multifactorial. The impacts of the clinical features such as age and the immune status in promoting SARS-CoV-2 infection and severe illness among CLL patients are acknowledged although the full picture is not yet shaped (9).

It has also remained to understand the influence of host genetic features related to SARS-CoV-2 infection which could potentially provoke the comorbidity of CLL and COVID-19 disease. The well accepted entry process for SARS-CoV-2 is that the virus binds through its spike (S) proteins to a cellular receptor, the SARS-CoV-2’s ‘S’ proteins is then cleaved by a cellular protease and fuse the receptor transmembrane for penetration into nucleus. In principle, the patterns of expression of certain genes related to the comorbid disease such as cancer disease could have an influence on the expression of host proteins which regulate viral entry into cells and therefore could promote or restrict SARS-CoV-2 infection (11). Several studies described a number of molecules linked with SARS-CoV-2 entry into host cells, in which some found higher gene expression of entry molecules in cancer diseases and suggested to be linked with severe COVID-19 disease observed in patients albeit findings mainly based on bioinformatic analysis (12-14). The gene expression profile of molecules related to cellular entry of SARS-CoV-2 in CLL disease are uncertain. Such information could help in delineating the molecular basis underling comorbidity of CLL disease among COVID-19 infected patients and may also provide rationales for therapeutic interventions to counteract or minimize the severity of COVID-19 disease in CLL patients. The aim of this study was to decipher the expression pattern of genes that may promote or restrict SARS-CoV-2 entry based on using of CLL and normal B cells microarray datasets.

2. MATERIALS AND METHODS:

I) Selection of SARS-CoV-2 entry-related gene list: The selection of cellular factors related to SARS-CoV-2 entry for analysis of gene expression in CLL disease was based on reported data from Singh et al. (2020) study which curated a list of genes through mining the literature in search for confirmed and candidate genes related to
SARS-CoV-2 entry into human cells. The study identified 28 genes, referred to as SARS-CoV-2 and coronavirus-related receptors and factors (SCARFs), for single-cell transcriptomic analysis based on several tissues derived from healthy humans (15). For the purpose of current study, the analysis of gene expression was limited to SCARFs which confirmed or suggested to work as receptors, mediators and restrictors of SARS-CoV-2 entry and therefore SCARFs act post SARS-CoV-2 infection (factors required for viral genome replication or traffic and/or assembly) were excluded. The SCARFs included in this study are 11 genes which promote SARS-CoV-2 entry as well as 4 candidates for restriction of SARS-CoV-2 and/or CoV entry and are as the following: 1- Receptors for SARS-CoV-2 and/or CoV (ACE2, BSG, DDP4, ANPEP, CD209, CLEC4G and CLEC4M), 2- Proteases mediate entry of SARS-CoV-2 and/or CoV (TMPRSS2, TMPRSS4, CTSB and Furin) and 3- Candidates for restriction of SARS-CoV-2 and/or CoV entry into cells (IFITM1, IFITM2, IFITM3 and LY6E).

II) Acquisition and processing of Microarray datasets:
The expression analysis of candidate genes which could promote or restrict SARS-CoV-2 entry into human cells was performed based on publicly available microarray datasets of peripheral blood samples from CLL patients and normal B cells (NBCs). The NCBI-Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/gds/) was queried for studies which focused on microarray analysis of gene expression between human CLL and Mature NBCs. A total of 6 studies were identified by 20 Oct. 2020 as candidate sources of microarray data for this study. The three studies which contained > 100 samples of CLL samples (GSE50006, GSE31048 and GSE46261) were selected for analysis of expression profiles of candidate genes in CLL patients (Table 1).

<table>
<thead>
<tr>
<th>Geo ID</th>
<th>NBCs</th>
<th>CLL</th>
<th>Total</th>
<th>Platforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE50006</td>
<td>32</td>
<td>188</td>
<td>220</td>
<td>Affymetrix Human Genome U133 Plus 2.0 Array</td>
</tr>
<tr>
<td>GSE31048</td>
<td>24</td>
<td>179</td>
<td>203</td>
<td>Affymetrix Human Genome U133 Plus 2.0 Array</td>
</tr>
<tr>
<td>GSE46261</td>
<td>6</td>
<td>211</td>
<td>217</td>
<td>Affymetrix Human Gene 1.0 ST Array</td>
</tr>
<tr>
<td>GSE26725</td>
<td>5</td>
<td>12</td>
<td>17</td>
<td>Affymetrix Human Genome U133 Plus 2.0 Array</td>
</tr>
<tr>
<td>GSE18026</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>Affymetrix Human Genome U133 Plus 2.0 Array</td>
</tr>
<tr>
<td>GSE36907</td>
<td>5</td>
<td>7</td>
<td>12</td>
<td>Affymetrix Human Genome U133 Plus 2.0 Array</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>and Affymetrix Human Gene 1.0 ST Array</td>
</tr>
</tbody>
</table>

III) Processing of Microarray datasets:
The unprocessed data files of individual dataset were retrieved from the NCBI-GEO database in the form of a “.cel” file and uploaded into the GenePattern software which is an open access suite providing computational analysis of genomic data (16). The “.cel” files of each dataset were normalized by Robust Multi-array Average (RMA) algorithm with quantile normalization and background correction and then processed for collapsing probes to genes based on the maximum expression value. The expression values of the genes of interest were extracted out of the total genes, subjected to log transformation (log-2) and then processed in
the Comparative Marker Selection module within the GenePattern to determine the differentially expressed genes (DEGs) between CLL and NBCs using the two-sided student’s t-test and 10000 permutations of phenotype labels. To perform the meta-Analysis, the log transformed files of the three datasets were merged into one dataset (CLL n=578 and NBCs n= 62; total=640) and submitted into the ComBat module, which use an Empirical Bayes approach, to correct for batch effects (17). The adjusted data was then processed to identify the DEGs between CLL and NBCs using the same statistical parameters used for the analysis of individual dataset.

IV) Statistical Analysis:
The statistical measurements for the differences in the average log-2 gene expression, (i.e. t-test, P value, Fold discovery rate (FDR) and Fold Change), between CLL and NBCs were calculated by the Comparative Marker Selection module within the GenePattern. The level of significance in DEGs between CLL and NBCs was considered based on P value; Non-significant difference in gene expression denote $p \geq 0.05$ and the significant difference in gene expression is indicated by $p<0.05$.

3. RESULTS:

I) Analysis the expression of SCARFs genes in CLL based on individual dataset:
The first analysis in this study focused on assessing the differential expression of SCARFs genes between CLL and NBCs based on individual microarray dataset. I sought to initially identify the pattern of the differential expression in the 15 SCARFs genes between CLL and NBCs, based on a cutoff $p$ value $< 0.05$, in each dataset individually and then compare the common and unique differential expression between the three studies (i.e. GSE50006, GSE31048 and GSE46261). As depicted in Figure 1, the patterns of the DEGs between CLL and NBCs were comparable between datasets generated using the same microarray platform (GSE50006 and GSE31048) whereas GSE46261, which generated using different platforms, demonstrated a relatively unique pattern of DEGs between CLL and NBCs. This initial analysis of the DEGs profiles indicated the necessity for removal of batch effect, which could be introduced from using different platforms, for interrogation of a more precise estimate of the DEGs between CLL and NBCs.
Figure 1: Venn diagrams showing the overlap in the differential expression of SCARFs genes between CLL datasets.

A) Insignificantly expressed genes in CLL datasets. B) Significantly over expressed genes in CLL datasets. C) Significantly under expressed genes in CLL datasets. The cutoff level of significance is \( P \text{ value} < 0.05 \). The table below graphs list common and unique differential expression of SCARFs genes between the three studies (i.e. GSE50006, GSE31048 and GSE46261).

II) Analysis the expression of SCARFs gene in CLL based on meta-analysis:
The main analysis of this study was based on compiling the datasets and then removing of batch effect using the empirical Bayes method (ComBat) for meta-analysis of difference in SCARFs genes between CLL and NBCs. The pattern of the differential expression of SCARFs genes between CLL and NBCs based on meta-analysis study is shown in Figure 2.

Table 2. Evaluation of the expression of ACE2 gene, which is confirmed to be an entry receptor for SARS-CoV-2 (18), revealed no significant difference in expression between CLL and NBCs (Figure 2A). The difference in expression of BSG gene, which is confirmed as another receptor for entry of SARS-CoV-2 into human cells (18), was significantly high in CLL as compared to NBCs (Figure 2A). The next analysis was conducted on receptors which is confirmed to enable entry of either SARS-CoV or MERS-CoV into human cells (ANPEP, CD209, CLEC4G, CLEC4M and DPP4) and therefore could facilitate SARS-CoV-2 entry. As shown in Figure 2B, the expression levels of ANPEP and CD209 were not significantly increased in CLL compared to NBCs. In contrast, there were significant high expressions of CLEC4G, CLEC4M and DPP4 genes in CLL patients (Figure 2B). I further profiled the expression levels of the serine proteases which mediate entry of SARS-CoV-2 and/ or other human CoV into human cells (TMPRSS2, TMPRSS4, CTSB and Furin). The data showed no increase significant difference in the expression of all the proteases between CLL and NBCs (Figure 2C).
Several pieces of evidence support the integral role of many Interferon Stimulated Genes (ISGs) in restriction of SARS-CoV-2 entry into human cells and may therefore contribute to prevention of SARS-CoV-2 infection. The analysis in this study focused on restriction factors that are known to prevent infection of cells with a wide range of enveloped RNA viruses, including SARS-CoV (i.e., IFITM1, IFITM2, and IFITM3) (19) or protect cells against the entry of SARS-CoV-2 (i.e., LY6E) (20). Both IFITM1 and IFITM3 demonstrated non-statistically significant differences in expression between CLL and NBCs whereas IFITM2 and LY6E were significantly down expressed in the CLL dataset (Figure 2D).

Figure 2: The gene expression of SCARFs genes in CLL compared to NBCs based on meta-analysis of CLL datasets.

A) Confirmed receptors for SARS-CoV-2 (ACE2 and BSG), B) Potential receptors for SARS-CoV-2 (DDP4, ANPEP, CD209, CLEC4G and CLEC4M), C) Protease for SARS-CoV-2 entry (TMPRSS2, TMPRSS4, CTSB and Furin) and D) Candidates for restriction of SARS-CoV-2 are (IFITM1, IFITM2, IFITM3 and LY6E). NS indicates non-significant (p≥0.05), * indicates p<0.05, ** indicates p<0.01, *** indicates p<0.001.

Table 2: The statistical measurements for the differences in expression of SCARFs genes between CLL and NBCs in the meta-analysis study.

<table>
<thead>
<tr>
<th>SCARFs Class</th>
<th>Feature</th>
<th>Description</th>
<th>t-test</th>
<th>P value</th>
<th>FDR*</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry-Receptors</td>
<td>ACE2</td>
<td>Angiotensin I converting enzyme (peptidyl-dipeptidase A) 2</td>
<td>1.44</td>
<td>0.151</td>
<td>0.247</td>
<td>1.011</td>
</tr>
<tr>
<td></td>
<td>BSG</td>
<td>Basigin (Ok blood group) and CD147</td>
<td>8.26</td>
<td>0.0002</td>
<td>0.001</td>
<td>1.060</td>
</tr>
<tr>
<td></td>
<td>ANPEP</td>
<td>Alanyl (membrane) aminopeptidase (aminopeptidase N, aminopeptidase M, microsomal aminopeptidase, CD13 and Gp150)</td>
<td>0.09</td>
<td>0.953</td>
<td>0.953</td>
<td>1.002</td>
</tr>
</tbody>
</table>
DISCUSSION:

Considering the serious illness and the increased mortality from COVID-19 disease observed in some cancer patients, it is been hypothesized that cancer cells might have higher expression of genes related to SARS-CoV-2 entry (21). Certainly, several studies supported this concept through the finding that molecules linked with the entry of SARS-CoV-2 into host cells are commonly aberrantly expressed in some cancer types, which being considered at high risk of severe COVID-19 illness, compared to healthy people (11, 12, 21, 22), suggesting for a possible host genetic factor for such risky associations. In the context of CLL, there is a lack of knowledge about the expression of factors which promote or restrict the entry of SARS-CoV-2 into CLL cells.

The current study aimed to investigate the expression of a list of genes, known as SCARFs in CLL disease based on integrative analysis of multiple microarray datasets. While interrogation of the DEGs based on large size of samples from multiple microarray datasets

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>log2 Fold Change</th>
<th>FDR</th>
<th>Q-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD209</td>
<td>CD209 molecule, DC-SIGN and CLEC4L</td>
<td>1.21</td>
<td>0.223</td>
<td>0.304</td>
<td>1.007</td>
</tr>
<tr>
<td>CLEC4G</td>
<td>C-type lectin superfamily 4 member G, LSECTin and DTTR431</td>
<td>9.28</td>
<td>0.0002</td>
<td>0.001</td>
<td>1.070</td>
</tr>
<tr>
<td>CLEC4M</td>
<td>C-type lectin domain family 4 member M, DC-SIGNR, LSIGN, CD299 and CD209L</td>
<td>9.82</td>
<td>0.0002</td>
<td>0.001</td>
<td>1.068</td>
</tr>
<tr>
<td>DPP4</td>
<td>Dipeptidyl-peptidase 4, CD26 and adenosine deaminase complexing protein 2</td>
<td>3.11</td>
<td>0.005</td>
<td>0.014</td>
<td>1.060</td>
</tr>
<tr>
<td>TMPRSS2</td>
<td>Transmembrane protease serine 2, Epitheliasin</td>
<td>0.64</td>
<td>0.508</td>
<td>0.586</td>
<td>1.005</td>
</tr>
<tr>
<td>TMPRSS4</td>
<td>Transmembrane protease serine 4, CAPH2</td>
<td>1.46</td>
<td>0.134</td>
<td>0.247</td>
<td>1.009</td>
</tr>
<tr>
<td>CTSB</td>
<td>Cathepsin B</td>
<td>0.98</td>
<td>0.372</td>
<td>0.465</td>
<td>1.016</td>
</tr>
<tr>
<td>FURIN</td>
<td>Paired basic amino acid cleaving enzyme</td>
<td>-2.53</td>
<td>0.011</td>
<td>0.028</td>
<td>1.029</td>
</tr>
<tr>
<td>IFITM1</td>
<td>Interferon induced transmembrane protein 1 (9-27)</td>
<td>-0.46</td>
<td>0.650</td>
<td>0.697</td>
<td>1.004</td>
</tr>
<tr>
<td>IFITM2</td>
<td>Interferon induced transmembrane protein 2 (1-8D)</td>
<td>-2.25</td>
<td>0.025</td>
<td>0.054</td>
<td>1.015</td>
</tr>
<tr>
<td>IFITM3</td>
<td>Interferon induced transmembrane protein 3 (1-8U)</td>
<td>-1.33</td>
<td>0.164</td>
<td>0.247</td>
<td>1.013</td>
</tr>
<tr>
<td>LY6E</td>
<td>Lymphocyte antigen 6 complex, locus E</td>
<td>-3.72</td>
<td>0.001</td>
<td>0.004</td>
<td>1.033</td>
</tr>
</tbody>
</table>
provide more precise estimate of the DEGs, the experimental variations and the difference in microarray platforms used to generate datasets could contribute to batch effect and therefore confound the robustness and accuracy of the DEGs (23). To overcome this limitation, the datasets were compiled and processed for removal of batch effect prior to conduct the meta-analysis study for analysing the differential expression of SCARFs genes in CLL.

The results from the meta-analysis study did not show considerable influence of CLL disease on the gene expression of the well-known entry receptor (i.e. ACE-2) and protease (i.e. TMPRSS2) for SARS-CoV-2. This finding suggests that the virus could use different entry receptor/s and co-receptors for cellular entry in CLL. Perhaps the most intriguing result in the meta-analysis study is the significant higher expression of genes which encode for other candidate receptors for SARS-CoV-2 (i.e. BSG, CLEC4G, CLEC4M and DDP4) in CLL. Evidence from the literature indicates that BSG and CLEC4M act as receptors for SARS-CoV-2 in human tissues, CLEC4G and DDP4 potentially interact with SARS-CoV-2's ‘S’ proteins for viral entry and BSG, CLEC4M and DPP4 are highly expressed in different cancer types being linked with severe COVID-19 disease (14, 24). The finding that CLL increases the expression of these genes imply an increase in availability of host proteins for SARS-CoV-2 entry and hence a high risk for contracting SARS-CoV-2 and severe COVID-19 disease in CLL.

A crucial event mediated by the innate immune system against viral invasion is the production and binding of interferon molecules to their respective receptors for induction of ISGs to prevent viral infection. Although ISGs are known as interferon inducible molecules, the basal expression level of ISGs is potentially an integral determinant factor for SARS-CoV-2 tropism (15). Experimental functional studies have shown that infection of SARS-CoV-2 to human cells can be restricted by ISGs (IFITM1, IFITM2, IFITM3 and LY6E) (19, 20). In this study, the expression levels of two ISGs (IFITM2 and LY6E) were significantly low in CLL which could imply a breach in the defence against viral invasion to CLL cells.

The current view about the impact of COVID-19 disease on CLL patients indicates that an unacceptably high proportion of patients are susceptible to severe COVID-19 disease and fatal outcome. For instance, data from a multicenter study which assessed the outcomes of COVID-19 in 198 patients with CLL reported high mortality rate in both watch-and-wait and treated CLL patients, indicating that CLL patients are at increased risk for death from COVID-19 disease regardless of the stage of the disease or treatment status (9). The aberrant expression of factors supporting SARS-CoV-2 entry into cells in CLL disease observed in this study could represent, at least in part, the molecular basis of the severe COVID-19 infection observed in CLL patients. While the current research is hypothesis-generating study based on a retrospective analysis of CLL datasets, the large sample size used in this study is reasonable to support the applicability of the findings and provide preliminary data for prospective research aiming to investigate the implications of aberration expression of viral entry molecules on the severity of COVID-19 disease using samples from CLL patients with ongoing COVID-19 disease.
5. CONCLUSION:
In conclusion, this study evaluated the gene expression patterns of SARS-CoV2 entry-regulating molecules as potential factors that contribute to the comorbidity in CLL patients with COVID-19 disease. The analysis identified significant up regulation of 4 SARS-CoV-2 entry-promoting factors and down regulation of 2 SARS-CoV-2 entry-restricting factors in CLL disease. The findings from this study support findings from others that SARS-CoV-2 entry-regulating factors are significantly differentially expressed in COVID-19 comorbidities, including cancers and warrants further investigations into the possible roles of the host genetic factors related to SARS-CoV-2 entry in the comorbidity between CLL and COVID-19 disease.

Acknowledgement: The author thanks Dr. Abdulraheem Alsharif for his insightful revision, comments and suggestion about this work.
Authors’ Contributions: AA designed the study, gathered and entered the data, drafted the manuscript, revised the manuscript and finalized it.
Consent for Publication: Not applicable.
Ethics Approval and Consent to Participate: Not applicable.
Conflicts of Interest: The author declares no conflicts of interest regarding the publication of this paper

6. REFERENCE:


