Running Title: Targeting SARS-CoV-2 entry

Title: Genomics-guided identification of potential modulators of SARS-CoV-2 entry proteases, TMPRSS2 and Cathepsins B/L

Kartikay Prasad[1], and Vijay Kumar[1]*

[1]Amity Institute of Neuropsychology & Neurosciences, Amity University, Noida, UP- 201303, India

**Address for Correspondence*

Vijay Kumar, Ph. D.
Assistant Professor
Amity Institute of Neuropsychology & Neurosciences,
Amity University, Noida, UP- 201313, India
E-mail: vkumar33@amity.edu
ORCID: 0000-0002-3621-5025

Abstract

The entry of SARS-CoV-2 into host cells requires the activation of its spike protein by host cell proteases. The serine protease, transmembrane serine protease 2 (TMPRSS2) and cysteine proteases, cathepsins B, L (CTSB/L) activate spike protein and enabling SARS-CoV-2 entry to the host cell through two completely different and independent pathways. Given that the uncertainty of how SARS-CoV-2 infects and kills, the need for a deep understanding of SARS-CoV-2 biology is imperative. Herein, we performed genomic-guided meta-analysis to identify upstream regulatory elements altering the expression of TMPRSS2 and CTSB/L genes. Further, drugs and medicinal compounds were identified based on their effects on gene expression signatures of the modulators of TMPRSS2 and CTSB/L genes. Using this strategy, estradiol and retinoic acid have been identified as putative SARS-CoV-2 alleviation agents. Further, we analysed drug-gene and gene-gene interaction network using 332 human targets of SARS-CoV-2 proteins. The network results indicate that out of 332 human proteins, estradiol interacts with 135 (41%) and retinoic acid interacts with 40 (12%) proteins. Interestingly, a combination of both estradiol and retinoic acid interacts with 153 (46%) of human proteins acting as SARS-CoV-2 targets and affect the functions of nearly all of the SARS-CoV-2 viral proteins, indicating the therapeutic benefits of drug combination therapy. Finally, molecular docking analysis suggest that both the drugs binds to TMPRSS2 and CTSL with the nanomolar to low micromolar affinity. This study, for the first time, reports the molecules like estradiol and retinoic acid as candidate drugs against both the host proteases, TMPRSS2 and CTSB/L. We here thus suggest that these antiviral drugs alone or in combination can simultaneously target both the entry pathways and thus can be considered as a potential treatment option for COVID-19.

Keywords: COVID-19, TMPRSS2, Cathepsins, gene set enrichment, estradiol, retinoic acid

1. Introduction

The recent outbreak of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory 2 syndrome coronavirus 2 (SARS-CoV-2) has affected more than 25 million peoples with 1.5 million deaths worldwide and now is considered as a top 10 leading cause of death for the year 2020. While the world eagerly waiting for effective vaccines, scientists and clinicians exploring "off-label" use of countless drugs and compounds may help to manage and treat COVID-19 infection [1].

SARS-CoV-2 enters the host cell through spike (S) protein which binds to angiotensin-converting enzyme 2 (ACE2) receptor present on the cell-membrane of host cells [1]. This binding is then followed by the cleavage of the S protein by the host transmembrane serine protease 2 (TMPRSS2) and then fusion to the host cell membrane [2]. Reportedly, furin protease is also likely involved in the SARS-CoV-2 infection process [3]. Also, many other studies indicate the involvement of endosomal pathway as the key entry for SARS-CoV [4, 5]. These studies showed that the virus enters the host cell via pH- and receptor-mediated endocytosis pathway. In this pathway, the lysosomal cathepsins, mainly cathepsin L (CTSL) and cathepsin B (CTSB) cleaves and activate the S protein which then fuse with host cells [6, 7]. Very recently, Zhao et al. [8] reported that SARS-CoV-2 upregulates the expression of *CTSL* both in vivo and in vitro, which in turn enhances pseudovirus infection in human cells. These studies largely suggest that CTSL may be considered as a drug-target to treat COVID-19 infection [9, 10].

Therefore, the simultaneous co-expression of the host proteases, TMPRSS2 and CTSB/L in SARS-CoV-2 infected cells may lead to higher risk for COVID-19 infection. To get more insights into the regulation of their expression, we have performed genomic screening to find out the upstream regulatory elements altering the expression of TMPRSS2, CTSB, and CTSL genes (**Figure 1**). The drugs and medicinal compounds were then identified based on their

ability to change the gene expression of these regulatory elements. Using this strategy, we have identified and evaluate estradiol and retinoic acid that could be repurposed to ameliorate the outcomes of COVID-19 infection. Further, we utilized the network-based approach similar to our previous studies [11, 12] to examine the drug-gene and gene-gene interactions between the drugs and 332 human targets of SARS-CoV-2 proteins. To the best of our knowledge, the network-based approach of targeting viral entry routes simultaneously through a single drug molecule or in combination is still an unexplored therapeutic approach for COVID-19.

2. Methodology

2.1 Gene set enrichment analyses (GSEA) analysis of TMPRSS2 and CTSB/L

In this genomic-screen based study, gene set enrichment analyses (GSEA) were carried out to identify the genes linked to TMPRSS2 and CTSB/L for the significant enrichment of different functional categories. GSEA were performed through Enrichr bioinformatics webserver (http://amp.pharm.mssm.edu/Enrichr), which allows the investigation of nearly 200,000 genes from more than 100 gene set libraries [13, 14]. The predicted genes and transcription factors were further utilized for drugs and ligands identification based on their effects on gene expression signatures of the regulators of TMPRSS2 and CTSB/L genes. The screening for enrichment was based on the "combined score" which is a product of the log of the p-value obtained from the Fisher's exact test and the z-score, which is deviation from the expected rank (i.e. combined score, c = log(p)*z,). Next, the GSEA results were validated through computational investigations of the gene expression profiles from the NCBI Gene Expression Omnibus (GEO) database.

2.2. Drug-gene interaction Network

Further, the drug-gene interaction network of the predicted drug with 332 human genes targets of SARS-CoV-2 proteins identified form affinity-based proteomics study [15] was constructed and studied. Cytoscape [16] was used for preparing drug-gene interaction network.

2.3. Molecular Docking

After identifying the Estradiol and Retinoic acid as target molecules for CTSB/L and TMPRSS2, the three-dimension structural files of CTSB (PDB: 1HUC), and CTSL (PDB: 4AXL) were downloaded from Protein Data Bank [17]. At present, no structure of TMPRSS2 has been reported. The TMPRSS2 structure was prepared using the I-TASSER server [18]. I-TASSER is a threading based hierarchal approach for predicting the structure and function of proteins by identifying the structural templates for the target protein from PDB database using multiple threading approach. The three-dimensional structures were energy minimized using Swiss PDB Viewer (SPDBV) tool to obtain the lowest and stable energy conformation state of the proteins [19]. AutoDock tool [20] was used to prepare the structure of CTSB/L and TMPRSS2 by adding the Kollman charges along with the hydrogen atoms to polar groups in the protein structure. Further, the mol file of Estradiol and Retinoic acid was downloaded from the DrugBank database [21]. The OpenBabel [22] software was used for converting the mol file into three-dimensional structural file. AutoDock-Vina [23] was used for molecular docking and the structurally blind screening method was opted. The grid spacing was set to 1Å and the exhaustiveness was 8. The docking results were screened for high binding affinity, furthermore all the possible docked conformation was generated and the thorough analysis of interaction between the target proteins and ligand molecules were done using PyMOL (https://pymol.org/2/) and Discovery studio visualization tool [24].

3. Results

3.1. GSEA of genomic features associated with TMPRSS2 and CTSB/L

We have used the Enrichr bioinformatics platform [14] to identify human genes involved in regulatory cross talks and affecting the expression level and functions of the TMPRSS2 and CTSB/L genes, thus potentially affecting SARS-CoV-2 infection. To this end, GSEA were performed using the TMPRSS2 and CTSB/L genes as baits applied to extensive spectrum of

genomic databases regarding the structural, functional, regulatory, and pathophysiological features associated to these genes. The top-scoring genes were further examined using the Gene GEO database (**Figure 1**).

Expression profiling and GSEA of TMPRSS2 and CTSB/L genes revealed ubiquitous patterns of expression across human tissues. TMPRSS2 is highly expressed in bladder, kidney, as well as gastrointestinal and respiratory tract tissues. The CTSB/L genes are in general more homogeneously expressed across different tissues with highest expression in thyroid, salivary gland, and adipose tissues. (**Supplementary Figure S1**). TMPRSS2 as compared to CTSB/L has very low expression levels in brain and blood tissues.

GSEA of the COVID-19 related gene sets revealed that both TMPRSS2 and CTSL belongs to SARS top 50 geneshot AutoRIF along with SARS 133 Literature-Associated Genes from Geneshot GeneRIF (**Supplementary Figure S2 and Table S1**). The different GEO records revealed both the upregulation and downregulation of TMPRSS2 and CTSB/L genes by SARS-CoV-2 infection. Moreover, GEO records also suggest that the profile of expression largely depends on the type of cells (**Table S1**).

The virus perturbations data sets among GEO records of upregulated genes in Enrichr identified the SARS-CoV infection challenge at 60 hr (GSE47960) as the most significantly enriched record (**Figure 2**), marked by the upregulated expression of TMPRSS2 gene in human airway epithelial cells. Whereas CTSB showed upregulation in SARS-CoV infection after 24 hr (GSE47962) in human airway epithelial cells. These results indicate that SARS-CoV triggers the increased expression of both CTSB and TMPRSS2 gene after 1 day and 2.5 days respectively, of the infection (**Figure 2A and Supplementary Table S1**). These findings were substantiated by the increased expression of TMPRSS2 and CTSB/L reported in the peripheral blood mononuclear cells (PBMCs) of patients with severe acute respiratory syndrome (**Figure 2 B,C,D and Supplementary figure S3**) [25].

GSEA identified common human disorders ranging from cancer to neurological diseases as enriched records through the upregulation of these genes (**Supplementary Figure S4**), which is consistent with the clinical observations of comorbidity associated with COVID-19 infection. Interesting to note that both seasonal and pandemic H1N1 influenza virus infection significantly increased the expression of TMPRSS2 and CTSL in human bronchial epithelial cells whereas, CTSB showed both increased and decreased expression (**Supplementary Figure S4**).

Gene Ontology (GO) analyses revealed non-overlapping records of TMPRSS2 and CTSB/L genes (Supplementary Figure S5). The common significantly enriched records for biological process are proteolysis, and peptidase activity (acting on L-amino acid peptides). The biological process for TMPRSS2 are significantly enriched in positive regulation of viral entry into host cell (GO:0046598) and protein autoprocessing (GO:0016540), whereas, CTSB/L are enriched in cellular response to thyroid hormone stimulus (GO:0097067) and cellular protein catabolic process (GO:0044257).

3.2. Identification of the transcription factors associated with the TMPRSS2 and CTSB/L genes

GSEA of the enriched records of transcription factors (TFs) using ChEA 2016 and ENCODE TF ChIP-seq 2015 databases revealed different TFs associated with the TMPRSS2 and CTSB/L genes (**Supplementary Figure S6**). Common TFs shared by both TMPRSS2 and CTSB/L genes are PPARG (PPARγ) and SOX17 according to ChEA 2016. There was no common TFs identified through ENCODE TF ChIP-seq 2015 shared by these genes. However, for TMPRSS2 and CTSB, we found that PPARA (PPARα), RUNX1, and YAP1 binds to these two genes (**Supplementary Figure S6**).

Next, the GSEA analysis was done to find out the involvement of PPARγ, PPARα, SOX17 RUNX1, and YAP1 in COVID-19 infection. The GSEA analysis reported the differential

regulation of these genes in the COVID-19 gene sets available in Enrichr platform (**Supplementary Table S2**). PPARα has been shown to be downregulated by SARS-CoV-2 infection in intestinal organoids (GSE149312), in pancreatic organoids, and in liver organoids (GSE151803). PPARγ has also been shown to be downregulated upon SARS-CoV-2 infection in intestinal organoid (GSE149312) and SARS-CoV infection in airway epithelium (GSE47961). In contrast, PPARγ has been found to be upregulated in COVID-19 patient PBMC and SARS-CoV infection in Vero E6 cells (GSE30589). Also, SOX17 and RUNX1 were upregulated in cardiomyocytes by SARS-CoV-2 infection (GSE150392). YAP1 also gets upregulated in COVID-19 patients BALF and SARS-CoV-2 infection in liver organoids (GSE151803). Interestingly, the expression of PPARγ showed upregulation while PPARα and SOX17 showed mixed responses in the PBMC of patients with severe acute respiratory syndrome (**Supplementary figure S7**).

We also looked the GEO database to gauge the effects of these transcription factors on the expression of TMPRSS2 and CTSB/L genes. The GEO analysis indicate that PPARγ deficiency increases the expression of TMPRSS2 in induced inflammatory bowel disease (**Supplementary Figure S8**). Moreover, dominant negative expression of PPARγ decreased the expression of TMPRSS2 and CTSB/L genes, thus suggesting the repressor effects of PPARγ. Also, PPARα depletion increased the expression of TMPRSS2 and CTSB/L genes (**Supplementary Figure S8**). The other transcription factor, SOX17 overexpression leads to the increased expression of TMPRSS2 and CTSB/L genes (**Supplementary Figure S8**) suggesting the activation effects of SOX17. Similarly, the depletion of RUNX1 in human and mouse cells showed different effects on the expression of TMPRSS2 and CTSB/L genes. RUNX1 knockdown decreased the expression of TMPRSS2 in human prostate cancer cell line, whereas increased the expression of CTSB/L genes (**Supplementary Figure S8**). In contrast, RUNX1 depletion increased TMPRSS2 and CTSL expression, while decreased the expression

of CTSB in mouse cells (**Supplementary Figure S8**). The summary of these observations is reported in **Figure 3A**.

3.3. Identification of putative repressors of the TMPRSS2 and CTSB/L expression

GSEA of genomic databases were carried out to find putative modulators of TMPRSS2 and CTSB/L genes. The ARCHS4 transcription factor co-expression analysis indicated that the enriched records were demonstrating different patterns of co-expression with either TMPRSS2 or CTSB/L genes. Therefore, we look for the individual GSEA profiles for TMPRSS2 and CTSB, and for TMPRSS2 and CTSL genes. The GSEA of the TF perturbations followed by the expression database and the GEO gene perturbations database focused on upregulated genes identified POU5F1 and AIRE as potential repressors of the expression of TMPRSS2 and CTSB; and TMPRSS2 and CTSL expression, respectively (**Supplementary Figure S9**). These outcomes were further substantiated by observations that POU5F1 (also known as Oct4) knockdown significantly increases the expression of TMPRSS2 and CTSB/L expression in mouse embryonic stem cells (**Supplementary Figure S10**). AIRE deficiency on the other hand increases the expression of TMPRSS2 and CTSB while decreases the expression of CTSL in mouse thymic epithelial cells (**Supplementary Figure S10**). Moreover, both POU5F1 and AIRE genes have been mentioned in SARS133 literature-associated genes from Geneshot GeneRIF and also reported as up-regulated genes in COVID-19 infected bronchoalveolar lavage from patients (**Supplementary Table S2**). The summary of the results is provided in **Figure 3B**.

3.4. GSEA identify Estradiol and Retinoic acid as potential drugs for mitigating COVID-19 infection

Further, GSEA of the drug perturbations and ligand perturbations from GEO database records of downregulated genes revealed estradiol and retinoic acid as the top significantly enriched candidates (**Supplementary Figure S11**). Estradiol seems to affect both TMPRSS2 and

CTSB/L expression, while retinoic acid appears to target TMPRSS2 and CTSB expression (**Supplementary Figure S11**). These observations, thus provide the initial evidences suggesting that both estradiol and retinoic acid could be considered as the candidates for drug repurposing against SARS-CoV-2 infection.

Manual curation of GEO data sets suggested that both estradiol and retinoic acid exert biological activities which leads to alleviate the SARS-CoV-2 infection. The administration of estradiol inhibit TMPRSS2, CTSB, and CTSL expression in human endothelial cells (**Supplementary Figure S12**). Also, the administration of retinoic acid, has resulted in significantly decreased expression of TMPRSS2 and CTSB/L genes in human endothelial cells (**Supplementary Figure S12**).

Consistent with these findings, the investigation of GEO records revealed that estradiol appears to modulate several genes involved in promoting COVID-19 infection. Estradiol upregulates s the expression of PPARγ, and PPARα, while downregulates the expression of SOX17 and RUNX1 genes (**Supplementary Figure S13**). Furthermore, the interrogation of GEO records revealed that all-trans retinoic acid increased the expression of PPARγ and PPARα, while decreased the expression of SOX17 in human peripheral blood monocytes (**Supplementary Figure S13**).

Estradiol upregulates the expression of POU5F1 in mouse prostate gland (GSE3630). Also, the time course expression analysis of POU5F1 in mouse uterus response to 17beta-estradiol showed the upregulation for almost 4 hr (**Supplementary Figure S14**). Estradiol also increases the expression of POU5F1 and AIRE in endometrium of Rhesus monkey (**Supplementary Figure S14**). Retinoic acid also increases the expression of POU5F1 and AIRE in CD4+ T cells from spleen/lymph nodes of mouse (**Supplementary Figure S14**). Also, retinoic acid showed mixed response for the expression of AIRE in human peripheral blood monocytes and MCF-7 breast cancer cells, suggesting different mechanisms of regulation. **Figure 4**

summarizes the overall effects of estradiol and retinoic acid on the modifiers as well as on TMPRSS2 and CTSB/L.

3.5. Drug-gene interactions network of Estradiol and Retinoic acid in SARS-CoV2-human interactome

An excellent SARS-CoV-2-human interactome identified 332 high-confidence human protein targets of the 27 SARS-CoV-2 viral proteins [15]. We here hypothesize that the drug candidates significantly altering the protein-protein interactions of human proteins target of SARS-CoV-2 may inhibit the viral growth and replication. Moreover, the strength of therapeutics interference of a drug could be assessed by the numbers of drug-gene interactions. Notably, estradiol interacts with 135 of 332 (41%) human protein targets of SARS-CoV-2, and potentially interfering 26 of 27 (96%) SARS-CoV-2 proteins (**Figure 5A**). Most of these human genes were found to target more than one SARS-CoV-2 genes. These 135 genes make almost 322 interactions with 26 of SARS-CoV-2 proteins with maximum interactions to ORF8 (40), ORF9C (30), M and NSP13 (26) protein (**Figure 5B**), indicating an extensive protein-protein interaction network that can be modulated significantly by estradiol. Similarly, retinoic acid interacts with 40 out of 332 (12%) human proteins targeting SARS-CoV-2 and interfering with the activities of 23 of 27 (85%) SARS-CoV-2 proteins (**Figure 6A**). These 40 human genes interact with SARS-CoV-2 proteins and make a total of 80 interactions (**Figure 6B**) with maximum interactions observed in case to ORF8 (13), M (9), ORF9C, NSP13 and NSP7 (7) protein. Thus, estradiol and retinoic acid manifests significant interference with the SARS-CoV-2-human interactome. Interestingly, a combination of both estradiol and retinoic acid interacts to149 of 332 (45%) human proteins and thus affect the functions of nearly all (26 of 27; 96%) SARS-CoV-2 proteins (**Supplementary Figure 15).**

3.6 Identification of potential miRNAs involved in the regulation of TMPRSS2 and CTSB/L

miRNAs can be utilized as an essential antiviral tool. To determine potential miRNAs that can directly regulate to TMPRSS2 and CTSB/L, TargetScan microRNA 2017 has been explored. The analysis yielded no common miRNAs against these three genes (**Supplementary Figure 14**). However, three miRNAs have been identified to potentially bind against TMPRSS2 and CTSL genes. Out of this, one of the human miRNAs, hsa-miR-379 increased upon infection with influenza A infection in dendritic cells and in peripheral blood of Parkinson's disease patients (**Supplementary Figure 14b**). This miRNA has also been predicted to bind ORF10 of SARS-CoV-2 [26] and might be used as a therapeutic strategy against SARS-CoV-2 infection by altering its attachment potential inside the host's system.

3.7 Molecular docking of estradiol and retinoic acid to TMPRSS2 and CTSB/L genes

Molecular docking studies have largely been employed to know the binding modes of ligand and receptor, and is generally used in drug discovery. Through the molecular docking, we here evaluated the binding of estradiol and retinoic acid to TMPRSS2 and CTSB/L genes. The proposed binding mode is presented in Figures 7-8, and the docking results based on the binding affinity has been represented in **Table 1**.

The interaction analysis of estradiol with TMPRSS2 reveal the formation of two hydrogen bonds at ASN177 and TYR180 positions along with thirteen vander waal bonds and one pi-sulfur and alkyl bond each (**Figure 7A, B**). The interaction of estradiol with CTSL reveal the formation of two hydrogen bonds at the position, GLY20 and GLY194 along with seven vander waal bonds and two alkyl bonds (**Figure 7C, D**). Interaction study of estradiol with CTSB showed one hydrogen bond at VAL5 position along with nine vander waal bond, one carbon hydrogen bond and three other bonds (**Figure 7E, F**). Retinoic acid binds with CTSB, CTSL and TMPRSS2 with binding affinity of -7.9 kcal/mol, -6.0 kcal/mol and -6.4 kcal/mol, respectively. Retinoic acid form one hydrogen bond to VAL39 with TMPRSS2 along with

thirteen vander waal bonds, one carbon-hydrogen bond and one alkyl and pi-sigma bonds each (**Figure 8A, B**). Similarly, retinoic acid form two hydrogen bonds at position TRP189 and ASN18 with CTSL along with nine vander waal bonds and one alkyl bond (**Figure 8C, D**). Interaction study of retinoic acid with CTSB showed one hydrogen bond at position HIS96 along with five vander waal interactions, one pi-sigma bond and five alkyl bonds (**Figure 8E, F**). The interaction study represents how stringently is the binding of estradiol and retinoic acid with the target proteins which ultimately helps to lock the ligand molecules in the binding pocket and thus effectively inhibits the target proteins.

4. Discussion

Peroxisome proliferator-activated receptors (PPARs) belongs to ligand-activated nuclear hormone receptors (NR) superfamily and recently emerged as key players of inflammation. The PPAR family has PPAR-α, PPAR-β/δ, and PPAR-γ that are encoded by distinct genes. PPAR-γ expression is repressed in inflammatory lungs of patients with severe COVID-19. The repression of PPAR-γ play a key role in the induction of cytokine storm of inflammatory monocytes/macrophages in the SARS-CoV-2-infected lung. It has also been showed that SARS-CoV-2 alters lipid metabolism in the lung epithelial cells by modulating the expression of PPARα, and thus contributes to lipotoxicity and respiratory problems [27]. Thus, downregulation of PPARs in COVID-19 may be considered as an important effector of pulmonary inflammation and the potential mechanism behind the pathogenesis of acute lung injury [28]. In this regard, the activation of PPARs may serve as an effective therapeutic strategy to decrease the inflammatory perturbations during SARS-CoV-2 infection. Recently, Ehrlich et al reported that the PPARα agonist, fenofibrate decreased the phospholipid accumulation within SARS-CoV-2 infected cells, and inhibited viral replication [27]. Fenofibrate has been shown to suppress the downregulation of PPARα activation caused by

inflammation, decrease the cytokine production by LPS or TNFα [29, 30], and improve fatty acid oxidation, thus preventing acute lung injury [28].

Moreover, the activation of PPARs require the heterodimerization with another nuclear receptor, the retinoid X receptor (RXR) [31], that are activated by endogenous 9-cis retinoic acid [32], indicating a protective role of retinoic acid in lung injury.

Retinoic acid has been involved in the regulation of transcription of over 500 genes [33]. The pulmonary, immunomodulatory, and antimicrobial functions of retinoic acid plays a crucial role in reducing viral diseases, including COVID-19 infection [34]. It has been involved in modifying the pathogenesis of acute respiratory distress syndrome (ARDS), regulating the production of IL1-β, and IL-1 receptor antagonist and the subsequent pulmonary access of neutrophils [35]. In addition, combination of retinoic acid with simvastatin shown to be involved in pulmonary regeneration and remodeling in animal studies [36]. Retinoic acid also plays crucial role in viral infections because of its involvement in the development of innate immunity against RNA virus through type-I interferon-facilitated mechanism (retinoic acid-inducible gene I, RIG-1), which helps in protection of bystander immune cells against a subsequent round of viral replication [37].

Additionally, TMPRSS2 is sensitive to Dihydrotestosterone (DHT), and its expression is increased in a dosage-dependent manner. Retinoic acid has been found to inhibit DHT, thus resulting in an inhibition of androgen receptor stimulation, and downregulating the expression of TMPRSS2, and thus reduce SARS-CoV-2 infection. In light of this, a randomized, phase II, placebo controlled clinical trial (ClinicalTrials.gov; Identifier: NCT04578236) for assessing the efficacy of aerosol combination therapy of 13- cis retinoic acid and captopril for treating COVID-19 patients via indirect inhibition of TMPRSS2 is currently on-going.

The susceptibility to SARS-CoV-2 infection is almost similar in both genders, but males have higher severity and mortality. It has been noted that TMPRSS2 is an androgen-dependent

protein, signifying that SARS-CoV-2 infection is probably androgen-mediated [38]. Estradiol is a female sex hormone and shown to be protective against multiple pathological complications ranging from ARDS, inflammation, autoimmune diseases, viral infections to neurological disorders. Because of such wide-ranging protective effects, estradiol might control SARS-CoV-2 infection by affecting renin-angiotensin-aldosterone system (RAAS), suppressing inflammatory storms, inducing anti-viral immune responses, and enhancing the virus degradation through upregulation of endolysosomal degradation pathways [39, 40]. In line with this, a clinical trial testing the effect of sex hormones (estrogen and testosterone) on COVID-19 outcomes (ClinicalTrials.gov; Identifier: NCT04359329).

Interestingly, a hypothetical bipartite combination consisting of estradiol and retinoic acid may alter expression of 244 of 332 (73%) human genes encoding SARS-CoV-2 targets and interfere with the functions of all but one SARS-CoV-2 viral proteins. Notably, estradiol and retinoic acid significantly modulates the PPI network of human and SARS-CoV-2 proteins and thus manifest better therapeutic benefits by targeting a large number of genes involved in SARS-CoV-2 infection.

5. Conclusions

The main aim of this study was to identify human genes involved in regulation of the expression and functions of the SARS-CoV-2 entry genes, TMPRSS2 and CTSB/L. These identified genes may act as activators and /or repressors of TMPRSS2 and/or CTSB/L and provide necessary information to build a model of genomic regulatory interactions observed during COVID-19 infection. A panel of existing drugs and ligands against these regulatory genes were then identified that could be considered for drug-repurposing to mitigate the outcomes of COVID-19. Two of the most promising candidate drugs, namely estradiol and retinoic acid, manifest gene expression-altering activities of regulatory genes and thus could act as antiviral agents. Our findings are in excellent agreement with recent studies reporting the significant COVID-

19 mitigation potential of estradiol and retinoic acid [40-42]. Interestingly, a hypothetical bipartite combination (estradiol and retinoic acid) indicated more robust effects on interactome of SARS-CoV-2 target human host genes compared to monotherapies. In conclusion, our data provides a meaningful rich resource for future investigations of SARS-CoV-2 pathogenesis and also possible combinatorial therapeutic approaches to target COVID-19.

Author Contributions

Conceptualization, V.K; experiments and graphics, K.P; writing, V.K, and K.P; final review and editing, V.K.

Acknowledgments

Authors sincerely thank the Amity University, Noida for providing facilities.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] P. Zhou, X.L. Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B. Li, C.L. Huang, H.D. Chen, J. Chen, Y. Luo, H. Guo, R.D. Jiang, M.Q. Liu, Y. Chen, X.R. Shen, X. Wang, X.S. Zheng, K. Zhao, Q.J. Chen, F. Deng, L.L. Liu, B. Yan, F.X. Zhan, Y.Y. Wang, G.F. Xiao, Z.L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature, 579 (2020) 270-273.

[2] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Kruger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.H. Wu, A. Nitsche, M.A. Muller, C. Drosten, S. Pohlmann, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor, Cell, 181 (2020) 271-280 e278.

[3] M. Hoffmann, H. Kleine-Weber, S. Pohlmann, A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells, Mol Cell, 78 (2020) 779-784 e775.

[4] H. Wang, P. Yang, K. Liu, F. Guo, Y. Zhang, G. Zhang, C. Jiang, SARS coronavirus entry into host cells through a novel clathrin- and caveolae-independent endocytic pathway, Cell Res, 18 (2008) 290-301.

[5] J.K. Millet, G.R. Whittaker, Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis, Virus Res, 202 (2015) 120-134.

[6] Y. Zhou, P. Vedantham, K. Lu, J. Agudelo, R. Carrion, Jr., J.W. Nunneley, D. Barnard, S. Pohlmann, J.H. McKerrow, A.R. Renslo, G. Simmons, Protease inhibitors targeting coronavirus and filovirus entry, Antiviral Res, 116 (2015) 76-84.

[7] X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, L. Guo, R. Guo, T. Chen, J. Hu, Z. Xiang, Z. Mu, X. Chen, J. Chen, K. Hu, Q. Jin, J. Wang, Z. Qian, Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV, Nat Commun, 11 (2020) 1620.

[8] Miao-Miao Zhao, Wei-Li Yang, Fang-Yuan Yang, Li Zhang, Weijin Huang, Wei Hou, Changfa Fan, Ronghua Jin, Yingmei Feng, Youchun Wang, J.-K. Yang., Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development., medRxiv, (2020).

[9] T. Liu, S. Luo, P. Libby, G.P. Shi, Cathepsin L-selective inhibitors: A potentially promising treatment for COVID-19 patients, Pharmacol Ther, 213 (2020) 107587.

[10] S.P. Smieszek, B.P. Przychodzen, M.H. Polymeropoulos, Amantadine disrupts lysosomal gene expression: A hypothesis for COVID19 treatment, Int J Antimicrob Agents, 55 (2020) 106004.

[11] K. Prasad, F. Khatoon, S. Rashid, N. Ali, A.F. AlAsmari, M.Z. Ahmed, A.S. Alqahtani, M.S. Alqahtani, V. Kumar, Targeting hub genes and pathways of innate immune response in COVID-19: A network biology perspective, Int J Biol Macromol, 163 (2020) 1-8.

[12] F. Khatoon, K. Prasad, V. Kumar, Neurological manifestations of COVID-19: Available evidences and a new paradigm to understand the complications J Neurovirol, (2020).

[13] E.Y. Chen, C.M. Tan, Y. Kou, Q. Duan, Z. Wang, G.V. Meirelles, N.R. Clark, A. Ma'ayan, Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool, BMC Bioinformatics, 14 (2013) 128.

[14] M.V. Kuleshov, M.R. Jones, A.D. Rouillard, N.F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S.L. Jenkins, K.M. Jagodnik, A. Lachmann, M.G. McDermott, C.D. Monteiro, G.W. Gundersen, A. Ma'ayan, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, Nucleic Acids Res, 44 (2016) W90-97.

[15] D.E. Gordon, G.M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K.M. White, M.J. O'Meara, V.V. Rezelj, J.Z. Guo, D.L. Swaney, T.A. Tummino, R. Huttenhain, R.M. Kaake, A.L. Richards, B. Tutuncuoglu, H. Foussard, J. Batra, K. Haas, M. Modak, M. Kim, P. Haas, B.J.

Polacco, H. Braberg, J.M. Fabius, M. Eckhardt, M. Soucheray, M.J. Bennett, M. Cakir, M.J. McGregor, Q. Li, B. Meyer, F. Roesch, T. Vallet, A. Mac Kain, L. Miorin, E. Moreno, Z.Z.C. Naing, Y. Zhou, S. Peng, Y. Shi, Z. Zhang, W. Shen, I.T. Kirby, J.E. Melnyk, J.S. Chorba, K. Lou, S.A. Dai, I. Barrio-Hernandez, D. Memon, C. Hernandez-Armenta, J. Lyu, C.J.P. Mathy, T. Perica, K.B. Pilla, S.J. Ganesan, D.J. Saltzberg, R. Rakesh, X. Liu, S.B. Rosenthal, L. Calviello, S. Venkataramanan, J. Liboy-Lugo, Y. Lin, X.P. Huang, Y. Liu, S.A. Wankowicz, M. Bohn, M. Safari, F.S. Ugur, C. Koh, N.S. Savar, Q.D. Tran, D. Shengjuler, S.J. Fletcher, M.C. O'Neal, Y. Cai, J.C.J. Chang, D.J. Broadhurst, S. Klippsten, P.P. Sharp, N.A. Wenzell, D. Kuzuoglu-Ozturk, H.Y. Wang, R. Trenker, J.M. Young, D.A. Cavero, J. Hiatt, T.L. Roth, U. Rathore, A. Subramanian, J. Noack, M. Hubert, R.M. Stroud, A.D. Frankel, O.S. Rosenberg, K.A. Verba, D.A. Agard, M. Ott, M. Emerman, N. Jura, M. von Zastrow, E. Verdin, A. Ashworth, O. Schwartz, C. d'Enfert, S. Mukherjee, M. Jacobson, H.S. Malik, D.G. Fujimori, T. Ideker, C.S. Craik, S.N. Floor, J.S. Fraser, J.D. Gross, A. Sali, B.L. Roth, D. Ruggero, J. Taunton, T. Kortemme, P. Beltrao, M. Vignuzzi, A. Garcia-Sastre, K.M. Shokat, B.K. Shoichet, N.J. Krogan, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing, Nature, 583 (2020) 459-468.

[16] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome research, 13 (2003) 2498-2504.

[17] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic acids research, 28 (2000) 235-242.

[18] J. Yang, Y. Zhang, I-TASSER server: new development for protein structure and function predictions, Nucleic acids research, 43 (2015) W174-W181.

[19] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling, electrophoresis, 18 (1997) 2714-2723.

[20] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, Journal of computational chemistry, 30 (2009) 2785-2791.

[21] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, DrugBank: a comprehensive resource for in silico drug discovery and exploration, Nucleic acids research, 34 (2006) D668-D672.

[22] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: An open chemical toolbox, Journal of cheminformatics, 3 (2011) 33.

[23] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, Journal of computational chemistry, 31 (2010) 455-461.

[24] D.S. Biovia, Discovery studio modeling environment, Release, 2017.

[25] R. Reghunathan, M. Jayapal, L.Y. Hsu, H.H. Chng, D. Tai, B.P. Leung, A.J. Melendez, Expression profile of immune response genes in patients with Severe Acute Respiratory Syndrome, BMC Immunol, 6 (2005) 2.

[26] M.D. Sacar Demirci, A. Adan, Computational analysis of microRNA-mediated interactions in SARS-CoV-2 infection, PeerJ, 8 (2020) e9369.

[27] A. Ehrlich, S. Uhl, K. Ioannidis, M. Hofree, B.R. tenOever, Y. Nahmias, The SARS-CoV-2 Transcriptional Metabolic Signature in Lung Epithelium. , Cell Metabolism, (2020).

[28] H. Cui, N. Xie, S. Banerjee, J. Ge, S. Guo, G. Liu, Impairment of Fatty Acid Oxidation in Alveolar Epithelial Cells Mediates Acute Lung Injury, Am J Respir Cell Mol Biol, 60 (2019) 167-178.

[29] J. Becker, C. Delayre-Orthez, N. Frossard, F. Pons, Regulation of peroxisome proliferator-activated receptor-alpha expression during lung inflammation, Pulm Pharmacol Ther, 21 (2008) 324-330.

[30] C. Delayre-Orthez, J. Becker, I. Guenon, V. Lagente, J. Auwerx, N. Frossard, F. Pons, PPARalpha downregulates airway inflammation induced by lipopolysaccharide in the mouse, Respir Res, 6 (2005) 91.
[31] K.S. Miyata, S.E. McCaw, S.L. Marcus, R.A. Rachubinski, J.P. Capone, The peroxisome proliferator-activated receptor interacts with the retinoid X receptor in vivo, Gene, 148 (1994) 327-330.
[32] D.J. Mangelsdorf, U. Borgmeyer, R.A. Heyman, J.Y. Zhou, E.S. Ong, A.E. Oro, A. Kakizuka, R.M. Evans, Characterization of three RXR genes that mediate the action of 9-cis retinoic acid, Genes Dev, 6 (1992) 329-344.
[33] P. Lefebvre, P.J. Martin, S. Flajollet, S. Dedieu, X. Billaut, B. Lefebvre, Transcriptional activities of retinoic acid receptors, Vitam Horm, 70 (2005) 199-264.
[34] R. Caccialanza, A. Laviano, F. Lobascio, E. Montagna, R. Bruno, S. Ludovisi, A.G. Corsico, A. Di Sabatino, M. Belliato, M. Calvi, I. Iacona, G. Grugnetti, E. Bonadeo, A. Muzzi, E. Cereda, Early nutritional supplementation in non-critically ill patients hospitalized for the 2019 novel coronavirus disease (COVID-19): Rationale and feasibility of a shared pragmatic protocol, Nutrition, 74 (2020) 110835.
[35] S. Hashimoto, S. Hayashi, S. Yoshida, K. Kujime, S. Maruoka, K. Matsumoto, Y. Gon, T. Koura, T. Horie, Retinoic acid differentially regulates interleukin-1beta and interleukin-1 receptor antagonist production by human alveolar macrophages, Leuk Res, 22 (1998) 1057-1061.
[36] C. Yang, X. Yang, J. Du, H. Wang, H. Li, L. Zeng, W. Gu, J. Jiang, Retinoic acid promotes the endogenous repair of lung stem/progenitor cells in combined with simvastatin after acute lung injury: a stereological analysis, Respir Res, 16 (2015) 140.
[37] C. Trottier, M. Colombo, K.K. Mann, W.H. Miller, Jr., B.J. Ward, Retinoids inhibit measles virus through a type I IFN-dependent bystander effect, FASEB J, 23 (2009) 3203-3212.
[38] C. Wenham, J. Smith, R. Morgan, COVID-19: the gendered impacts of the outbreak, Lancet, 395 (2020) 846-848.
[39] N. Khan, Possible protective role of 17beta-estradiol against COVID-19, J Allergy Infect Dis, 1 (2020) 38-48.
[40] G.V. Glinsky, Tripartite Combination of Candidate Pandemic Mitigation Agents: Vitamin D, Quercetin, and Estradiol Manifest Properties of Medicinal Agents for Targeted Mitigation of the COVID-19 Pandemic Defined by Genomics-Guided Tracing of SARS-CoV-2 Targets in Human Cells, Biomedicines, 8 (2020).
[41] U. Seeland, F. Coluzzi, M. Simmaco, C. Mura, P.E. Bourne, M. Heiland, R. Preissner, S. Preissner, Evidence for treatment with estradiol for women with SARS-CoV-2 infection, BMC Med, 18 (2020) 369.
[42] A.R. Sarohan, COVID-19: Endogenous Retinoic Acid Theory and Retinoic Acid Depletion Syndrome, Med Hypotheses, 144 (2020) 110250.

Table 1. Molecular docking results of estradiol and retinoic acid to target TMPRSS2 and CTSB/L.

Protein-Drug	Binding energy (kcal/mol)	H-bond	Vander Waal bond	Other bonds
TMPRSS2-Estradiol	-7.4	ASN177, TYR180	VAL171, PRO170, HIS169, ASP134, TRP168, ASN131, GLN173, ASN193, ALA235, LYS234, SER233, TRP176, ASP175	TRP132, CYS172
CTSL-Estradiol	-6.8	GLY20, GLY194	GLN21, ASN18, TRP193, GLU192, LYS147, TYR146, PHE145	TRP189, LEU144
CTSB-Estradiol	-7.4	VAL5	GLY43, GLU4, GLY42, SER6, PRO57, ILE56, CYS70, CYS59, PRO58, SER171	ALA7, TYR54, LEU44, HIS61
TMPRSS2-Retinoic	-6.4	VAL39	ALA216, GLY217, ASN192, SER215, SRG55, THR214, GLU16, TYR190, HIS40, PRO41, ASN17, GLU38, HIS18	MET1, LEU3, TYR15
CTSB-Retinoic	-7.9	HIS96	LEU10, GLU122, PRO202, ARG203, THR204,	VAL3, VAL5, ALA30, TRP31, TRP34, LEU39
CTSL-Retinoic	-6.0	ASN18, TRP189	ASP162, HIS163, GLN19, ASN66, CYS65, GLY23, CYS22, GLN21, GLY20	CYS25

Figure Legends:

Figure 1. A strategic workflow describing the identification of candidate drugs targeting the regulators TMPRSS2 and CTSB/L identified through GSEA.

Figure 2. Effects of SARS-CoV-2 challenges on the expression of TMPRSS2 and CTSB/L genes. (A) GSEA of the virus perturbations from GEO focused on upregulated genes. Star in the figure represents the SARS-CoV-2 infection at 60 hr and 24 hr for TMPRSS2 and CTSB/L genes, respectively. GEO profiles of (B) TMPRSS2, (C) CTSB and (D) CTSL expression in peripheral blood mononuclear cells (PBMCs) of patients with SARS.

Figure 3. Upstream regulatory elements affecting the expression of TMPRSS2 and CTSB/L. (A) Figure represents PPARγ and PPARα as repressor while SOX17 as activators of TMPRSS2 and CTSB/L gene expression. On the other hand, RUNX1 acts both as activator and repressor of TMPRSS2 and CTSB/L genes. (B) Two transcription factors, POU5F1 and AIRE downregulates the gene expression of TMPRSS2 and CTSB/L.

Figure 4. Summarized results of GSEA and GEO analysis. Effects of estradiol and retinoic acid on modulators of TMPRSS2 and CTSB/L expression. Both estradiol and retinoic acid upregulates the expression of PPARγ and PPARα, POU5F1 and AIRE, while downregulates the expression of SOX17 and RUNX1. Also, estradiol and retinoic acid downregulates the expression of TMPRSS2 and CTSB/L genes.

Figure 5. Effects of estradiol on the drug-gene interactions network of SARS-CoV-2-human interactome. (A) Estradiol interacts with 135 (i.e ~40%) of 332 human proteins targeted by SARs-CoV-2, making 322 interactions in total, thus potentially interfering the functions of 26 of 27 of SARs-CoV-2 viral protein in human cells. (B) Interaction network of estradiol target human host genes and SARS-CoV-2 viral proteins. Total number of interactions shown were 322. Green color represents SARS-CoV-2 proteins and blue color represents human genes (prey of viral proteins) interacting with estradiol.

Figure 6. Effects of retinoic acid on the drug-gene interactions network of SARS-CoV-2-human interactome. (A) Out of 332 SARs-CoV-2 human target proteins, retinoic acid interacts with 40 (i.e ~24%) proteins, making 85 interactions in total, potentially interfering with functions of 23 of 27 (i.e 85%) of SARs-CoV-2 protein in human cells. (B) Interaction network of the 40 retinoic acid target human genes with 23 SARS-CoV-2 proteins. Total number of interactions shown were 85. Green color represents SARS-CoV-2 proteins and blue color represents retinoic acid targeted human genes.

Figure 7. Molecular docking interactions of estradiol to TMPRSS2, CTSB, and CTSL. Two-dimensional (2D) diagrams of TMPRSS2-estradiol interactions using Ligplot+ (A). The protein residues and interactions are colored accordingly and provided in figure. The potential binding poses in the three-dimensional structure of the protein (B). The black dotted line represents intermolecular hydrogen bond interactions. Similarly, 2D-diagrams of CTSB-estradiol interactions (C) and binding poses in 3-D structure (D) are shown. (E) and (F) represents 2-D interactions for CTSL-estradiol and 3-D binding poses, respectively.

Figure 8. Molecular docking interactions of retinoic acid to TMPRSS2, CTSB, and CTSL. Two-dimensional (2D) diagrams of (A) TMPRSS2- retinoic acid, (C) CTSB-retinoic acid, and (E) CTSL-retinoic acid interactions using Ligplot+. The protein residues and interactions are colored accordingly and provided in figure. The potential binding poses in the three-dimensional structure of the protein are represented in (B), (D), and (F).