Single-cell gene regulatory network analysis reveals potential mechanisms of action of antimalarials against SARS-CoV-2

James J. Cai[1,2,3,*], Daniel Osorio[1]

[1]Department of Veterinary Integrative Biosciences, [2]Department of Electrical and Computer Engineering, [3]Interdisciplinary Program of Genetics, Texas A&M University, College Station, TX 77843, USA.

ORCIDs: 0000-0002-8081-6725 (J.J.C.), 0000-0003-4424-8422 (D.O.)

[*]Corresponding author: J.J.C. (jcai@tamu.edu)

Abstract

The efficiency of antimalarials, chloroquine (CQ) and hydroxychloroquine (HCQ), in the prevention and treatment of coronavirus disease 2019 (COVID-19) is under intense debate. The mechanisms of action of antimalarials against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have not been fully elucidated. Here, we applied a network-based comparative analysis, implanted in our machine learning workflow—scTenifoldNet, to scRNA-seq data from COVID-19 patients with different levels of severity. We found that genes of the *Malaria* pathway expressed in macrophages are significantly differentially regulated between patients with moderate and severe symptoms. Our findings help reveal the mechanisms of action of CQ and HCQ during SARS-CoV-2 infection, providing new evidence to support the use of these antimalarial drugs in the treatment of COVID-19, especially for patients who are mildly affected or in the early stage of the infection.

Main Text

The efficiency of chloroquine (CQ) and hydroxychloroquine (HCQ) in the prevention and treatment of coronavirus disease 2019 (COVID-19) is under intense debate [1-5]. HCQ is the hydroxylated derivative of CQ; both are weak bases with a common flat aromatic core structure and proven antimalarial drugs. HCQ is also widely used as an immunomodulator to treat autoimmune diseases, especially systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA) [6]. CQ and HCQ are considered potent candidates to treat infection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)—the etiological agent of the COVID-19 [7, 8]. Experimental studies suggest that CQ and HCQ have the capability of inhibiting the replication of several intracellular micro-organisms [9], including SARS-CoV-2 *in vitro* [10, 11]. Several human studies have been conducted with both these drugs in COVID-19, and have shown significant improvement in some parameters in patients with COVID-19 [3, 12]. Although the use of CQ and HCQ in COVID-19 treatment has been recommended, the link between the mechanism of action of antimalarials and the mechanism of SARS-CoV-2 cellular infection is still missing. Without the link being established, how CQ and HCQ act against SARS-CoV-2 will remain elusive.

To this end, we used a newly published single-cell RNA sequencing (scRNA-seq) data set [13] to perform a transcriptomic comparative analysis between COVID-19 patients with moderate and severe symptoms.

The scRNA-seq data was collected from cells in bronchoalveolar lavage fluid (BALF) from three patients with moderate and six patients with severe infection symptoms [13]. We downloaded the raw data from the Sequence Read Archive (SRA) database using accession number SRP250732 and processed the data to generate scRNA-seq expression matrices. In the scRNA-seq data analysis, cells from patients with moderate (M) and severe (S) symptoms were pooled into M and S groups for comparison. After performing the data quality control and cell clustering, we extracted 1,125 and 3,735 macrophages from cells of M and S groups, respectively (**Fig. 1a**). The identity of macrophages was confirmed with marker genes including *CD68* (**Fig. 1b**). We chose to focus on macrophages due to the abundance of cells in the samples (77.4% of cells are macrophages in the data) and the importance of lung macrophages contributing to local inflammation, including recruiting inflammatory monocytic cells and neutrophils and attracting T cells, as suggested in the original paper [13]. Toll-like receptor gene, *TLR2*, was found to be highly expressed among macrophages in the S group (**Fig. 1c**).

To systematically compare macrophage transcriptomes between M and S patient groups, we employed a machine learning workflow—named scTenifoldNet (**Fig. 1d**), which we developed to construct and compare single-cell gene regulatory networks (scGRNs) [14]. The most important advantage of this network-based comparative analytical framework is its sensitivity. We have shown that scTenifoldNet can detect differential regulatory patterns between highly similar scRNA-seq samples to reveal gene regulatory changes, which are undetectable otherwise [14]. scTenifoldNet is built upon several machine learning methods including principal component regression, low-rank tensor approximation, and manifold alignment (**Fig. 1d**). We anticipated that, when applying to the scRNA-seq data of BALF from COVID-19 patients, scTenifoldNet could help decipher the molecular complexity of the data to achieve breakthroughs through constructing and comparing scGRNs in this cellular system affected by SARS-CoV-2. Indeed, scTenifoldNet took two expression matrices of M and S groups (containing 3,792 genes) as input and identified 27 highly differentially regulated genes [false discovery rate (FDR) < 0.05]: *BCAS4, ENO3, AD000091.1, JSRP1, SNX32, ATG10, NOP16, CR759762.1, SMPD4, BEST1, KCP, CR388220.1, DNAJC17,* **ABCB9***, EME2, KIFC1, AL671277.1, UBXN11, CR759790.1, AL669813.2, BX005428.2, BOLA2,* **PPT1***, SH3D19, OR7C1, MFSD11,* and *FKBP15* (**Fig. 1e**, genes are sorted according to the P-value significance; *PPT1* and *ABCB9* shown in bold are genes found in the *Lysosome* pathway of KEGG database). scTenifoldNet also generated a ranked list of genes (**Supplementary Data 1**), including the 27 highly significant genes and all the rest of genes, sorted according to the significance of the scTenifoldNet test. We analyzed the ranked gene list using pre-ranked Gene Set Enrichment Analysis (GSEA), which is a method for determining whether any prior gene sets show statistical significance with respect to a ranked gene list. The results of GSEA analysis showed that eight pathway gene sets are highly significant (adjusted P-value < 0.05, **Table 1**). Among the eight, the most informative one is the *Malaria* pathway from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (**Fig. 1e**, inset). Thus, scTenifoldNet analysis suggested that genes in the Malaria pathway are significantly differentially regulated between M and S patient groups. We mapped expressed genes in our data set to the KEGG Malaria pathway and highlighted genes highly expressed in severe patients with red and lowly expressed with green (**Fig. 1f**). In the map, we found that *TLR2/4* and *TGFB*—two genes on the pathway leading to immunosuppression, and *CXCL8*, *MCP1,* and *IL1A*—genes known to participate in the development of fever and metabolic acidosis [15], are highly expressed in severely affected patients. Upregulated expression of *TLR2* in immune cells is also reported in patients affected with SLE [16], for which HCQ is a drug for treatment.

Given that our machine learning workflow identified that the *Malaria* pathway-related gene expression program is associated with different COVID-19 symptom severity, it is easy to conjure up the idea that antimalarials are potential candidates for treating early SARS-CoV-2 infection, especially when the symptom is less severe. Our reasoning is further supported by mechanisms of action of CQ and HCQ—modulate cellular immunity by suppressing immune cell function and reducing the secretion of pro-inflammation cytokines [6, 17]. This idea is also supported by several other pathway gene sets identified by scTenifoldNet, which are differentially regulated between M and S groups. These pathways include *Cellular response to type I interferon (GO:0071357), Type I interferon signaling pathway (GO:0060337), Antigen processing and presentation, Interferon alpha/beta signaling Homo sapiens R-HSA-909733*, and *Chemokine-mediated signaling pathway (GO:0070098)* (**Table 1**).

The significant pathway gene sets also include *Phagosome* and *Endosomal/Vacuolar pathway Homo sapiens R-HSA-1236977* (**Table 1**). These pathway gene sets, along with *PPT1* and *ABCB9*, the two highly significant genes that are also in the lysosome pathway, suggest that different levels of symptom severity between M and S groups may be associated with the lysosome function. These results further justify the use of CQ and HCQ for COVID-19 treatment. An important mode of action of CQ and HCQ is the inhibition of lysosomal activity [18]. CQ and HCQ are weak bases that are known to elevate the pH of acidic intracellular organelles, such as endosomes and lysosomes, essential for membrane fusion [19]. If SARS-CoV-2 is internalized by receptor-mediated endocytosis and delivered to lysosomes, the virus can be prevented by CQ and HCQ that block function of the lysosome. In the presence of CQ and HCQ, SARS-CoV-2 viruses that require acidic pH to fuse with the cell membrane can no longer do so, and thus cells are protected from infection.

In conclusion, we applied the network-based comparative analysis, implanted in our machine learning workflow scTenifoldNet, to scRNA-seq data from COVID-19 patients with different levels of severity. We found that genes of the *Malaria* pathway expressed in macrophages are significantly differentially regulated between patients with moderate and severe symptoms. Our findings help reveal the mechanisms of action of CQ and HCQ during SARS-CoV-2 infection, providing new evidence to support the use of these antimalarial drugs in the treatment of COVID-19, especially for patients who are mildly affected or in the early stage of the infection. The code and data of our study are available at https://github.com/cailab-tamu/covid19-antimalarials-letter.

References

1. Joob, B. and V. Wiwanitkit, *Evidence of Protective Effect of Hydroxychloroquine on COVID-19.* J Rheumatol, 2020.
2. Romao, V.C., A.R. Cruz-Machado, and J.E. Fonseca, *No evidence so far on the protective effect of hydroxychloroquine to prevent COVID-19: response to the comment by Joob and Wiwanitkit.* Ann Rheum Dis, 2020.
3. Gautret, P., et al., *Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial.* Int J Antimicrob Agents, 2020: p. 105949.
4. Toumi, M. and S. Aballea, *Commentary on "Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open label non-randomized clinical trial" by Gautret et al.* J Mark Access Health Policy, 2020. **8**(1): p. 1758390.
5. Arshad, S., et al., *Treatment with Hydroxychloroquine, Azithromycin, and Combination in Patients Hospitalized with COVID-19.* Int J Infect Dis, 2020.

6.   Schrezenmeier, E. and T. Dorner, *Mechanisms of action of hydroxychloroquine and chloroquine: implications for rheumatology.* Nat Rev Rheumatol, 2020. **16**(3): p. 155-166.

7.   Liu, J., et al., *Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro.* Cell Discov, 2020. **6**: p. 16.

8.   Colson, P., et al., *Chloroquine and hydroxychloroquine as available weapons to fight COVID-19.* Int J Antimicrob Agents, 2020. **55**(4): p. 105932.

9.   Rolain, J.M., P. Colson, and D. Raoult, *Recycling of chloroquine and its hydroxyl analogue to face bacterial, fungal and viral infections in the 21st century.* Int J Antimicrob Agents, 2007. **30**(4): p. 297-308.

10.   Yao, X., et al., *In Vitro Antiviral Activity and Projection of Optimized Dosing Design of Hydroxychloroquine for the Treatment of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2).* Clin Infect Dis, 2020.

11.   Wang, M., et al., *Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro.* Cell Res, 2020. **30**(3): p. 269-271.

12.   Gao, J., Z. Tian, and X. Yang, *Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies.* Biosci Trends, 2020. **14**(1): p. 72-73.

13.   Liao, M., et al., *Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19.* Nat Med, 2020. **26**(6): p. 842-844.

14.   Osorio, D., et al., *scTenifoldNet: a machine learning workflow for constructing and comparing transcriptome-wide gene regulatory networks from single-cell data.* bioRxiv, 2020: p. 2020.02.12.931469.

15.   Dinarello, C.A., *The history of fever, leukocytic pyrogen and interleukin-1.* Temperature (Austin), 2015. **2**(1): p. 8-16.

16.   Liu, Y., et al., *Increased expression of TLR2 in CD4(+) T cells from SLE patients enhances immune reactivity and promotes IL-17 expression through histone modifications.* Eur J Immunol, 2015. **45**(9): p. 2683-93.

17.   Fox, R., *Anti-malarial drugs: possible mechanisms of action in autoimmune disease and prospects for drug development.* Lupus, 1996. **5 Suppl 1**: p. S4-10.

18.   Gies, V., et al., *Beyond Anti-viral Effects of Chloroquine/Hydroxychloroquine.* Frontiers in Immunology, 2020. **11**(1409).

19.   Mauthe, M., et al., *Chloroquine inhibits autophagic flux by decreasing autophagosome-lysosome fusion.* Autophagy, 2018. **14**(8): p. 1435-1455.

Table 1. Significant pathway gene sets identified using scTenifoldNet-ranked genes with GSEA analysis.

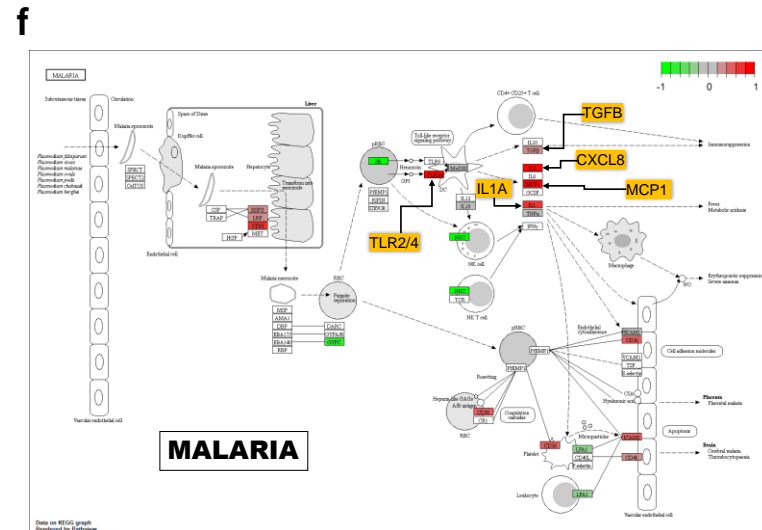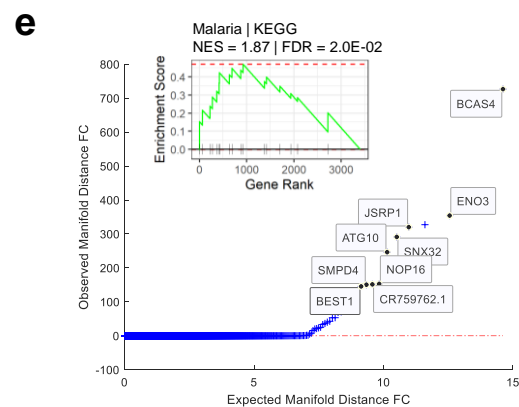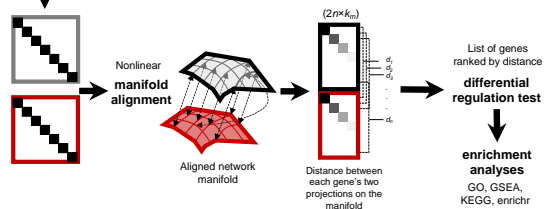|	Pathway gene set	P-value	Adjusted P-value	Size of gene set	Representative genes (i.e., leading edges of GSEA)
1	Cellular response to type I interferon (GO:0071357)	8.62E-07	0.002	37	*MX1, STAT1, ADAR, SAMHD1, OASL, IFIT2, HLA-C, ISG15, HLA-A, HLA-B, IFI6, SP100, RSAD2, IFI27, IFITM3, OAS3, IFIT1, IFIT3, OAS1, MYD88, PSMB8, GBP2, OAS2*
2	Type I interferon signaling pathway (GO:0060337)	8.62E-07	0.002	37	*MX1, STAT1, ADAR, SAMHD1, OASL, IFIT2, HLA-C, ISG15, HLA-A, HLA-B, IFI6, SP100, RSAD2, IFI27, IFITM3, OAS3, IFIT1, IFIT3, OAS1, MYD88, PSMB8, GBP2, OAS2*
3	Antigen processing and presentation	2.34E-05	0.02	35	*IFI30, CD74, CTSB, HLA-DQA1, CTSL, HLA-C, TAP1, HLA-A, HLA-B, PDIA3, HSP90AB1, HSPA5, B2M, LGMN, HSPA8, CTSS, HLA-DRB1, HSPA1B, CALR*
4	Malaria	1.57E-05	0.02	14	*CCL2, TLR2, LRP1, PECAM1, MYD88, CD36, THBS1, THBS3, IL1B, ITGB2, ITGAL*
5	Phagosome	8.57E-05	0.02	71	*ACTB, HLA-DQA1, CTSL, HLA-C, TAP1, HLA-A, HLA-B, TLR2, EEA1, MARCO, RAB5C, FCGR2A, TUBB, CYBB, FCGR3A, CD14, CD36, THBS1, ATP6V1C2, M6PR, THBS3, ATP6V1C1, CTSS, ITGB2, HLA-DRB1, CYBA, ATP6V1B2, HGS, CALR, DYNC1H1, CORO1A, ATP6AP1, CLEC7A, ACTG1, FCGR2C, HLA-E, RAB5B, NCF2, MSR1, TLR4, SEC22B, TUBA1C, HLA-DQB1, TUBB4B*
6	Endosomal/Vacuolar pathway Homo sapiens R-HSA-1236977	4.55E-05	0.03	9	*CTSL, HLA-C, HLA-A, HLA-B, LNPEP, B2M, CTSS*
7	Interferon alpha/beta signaling Homo sapiens R-HSA-909733	2.61E-05	0.03	38	*MX1, STAT1, ADAR, SAMHD1, OASL, IFIT2, HLA-C, ISG15, HLA-A, HLA-B, IFI6, RSAD2, IFI27, IFITM3, OAS3, IFIT1, IFIT3, OAS1*
8	Chemokine-mediated signaling pathway (GO:0070098)	2.85E-05	0.04	9	*CCL2, CXCL10, CCL8, CCL7, CXCL11*

a

Moderate (n=1125)
Severe (n=3735)

t-SNE 2 / t-SNE 1

b CD68

c TLR2

d

scRNA-seq data → Random or pseudotime-guided **cell subsampling** → ($n \times m'$, $m' < m$) subsamples → Network construction **PC regression** → Adjacency matrices ($n \times n \times t$) t networks → CANDECOMP/PARAFAC (CP) **tensor decomposition** → Denoised adjacency matrices ($n \times n \times t$) t networks → Denoised weight-averaged network ($n \times n$)

($n \times m$ gene-cell matrix)

$c_1 \quad c_2 \quad c_t$
$b_1 + b_2 + \ldots + b_t$
$a_1 \quad a_2 \quad a_t$

Nonlinear **manifold alignment** → Aligned network manifold → ($2n \times k_m$) → List of genes ranked by distance **differential regulation test** → **enrichment analyses** GO, GSEA, KEGG, enrichr

Distance between each gene's two projections on the manifold

f

MALARIA

TGFB
CXCL8
IL1A
MCP1
TLR2/4

MALARIA

Data on KEGG graph
Rendered by Pathview

e

Malaria | KEGG
NES = 1.87 | FDR = 2.0E-02

Enrichment Score / Gene Rank

BCAS4
ENO3
JSRP1
ATG10
SNX32
SMPD4
NOP16
BEST1
CR759762.1

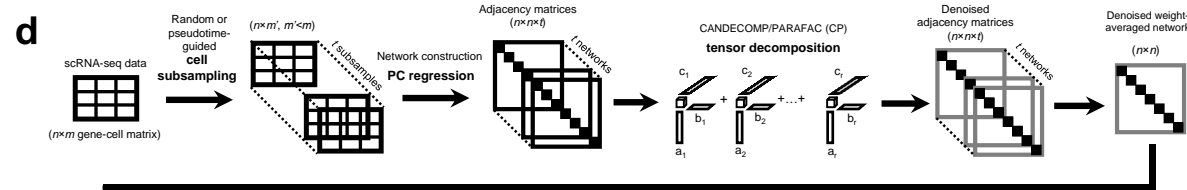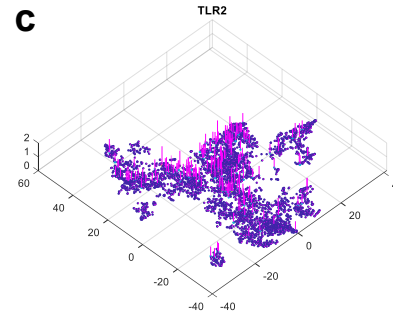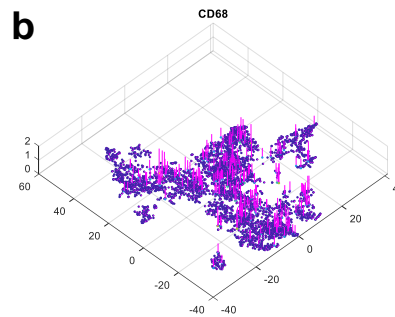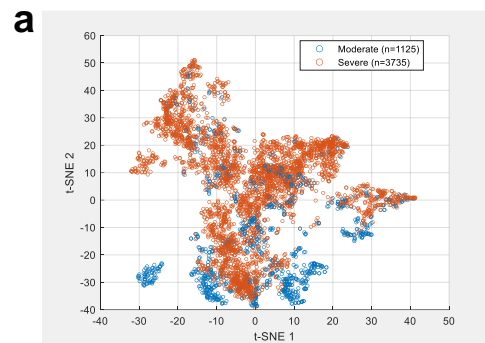Observed Manifold Distance FC / Expected Manifold Distance FC

Fig. 1. Expression of genes in the Malaria pathway is differentially regulated in macrophages from COVID-19 patients with moderate and severe symptoms. (**a**) t-distributed stochastic neighbor embedding (t-SNE) plot of macrophages used in the data analysis. Cells are colored according to the group of COVID-19 patients: moderate (blue) and severe (red). (**b**) Gene expression level of CD68 indicated with bars (magenta) on t-SNE plot of cells. (**c**) Gene expression level of *TLR2*. (**d**) scTenifoldNet workflow. scTenifoldNet is a machine learning framework that uses a comparative network approach with scRNA-seq data to identify regulatory changes between samples. scTenifoldNet is composed of five major steps. 1. *Cell subsampling*. scTenifoldNet starts with subsampling cells in the scRNA-seq expression matrices. Cells are subsampled either randomly or following a pseudotime trajectory of cells. The subsampling is repeated multiple times to create a series of subsampled cell populations. 2. *Network construction*. The subsampled data matrices are subject to network construction and form a multilayer single-cell gene regulatory network (scGRN). Principal component regression is used for network construction; each scGRN is represented as a weighted adjacency matrix. 3. *Tensor denoising*. The multilayer scGRN constructed from the subsampled data matrices is treated as a three-order tensor, which is subsequently decomposed into multiple components. Top components of tensor decomposition are then used to reconstruct denoised multilayer scGRN. The denoised multilayer scGRN is collapsed by taking average weight across layers. 4. *Manifold alignment*. Two denoised scGRNs: one from the first sample and the other from the other sample to be compared, are aligned with respect to common genes using a nonlinear manifold alignment algorithm. Each gene is projected to a low-rank manifold space as two data points, one from each sample. 5. *Differential regulation test*. The distance between the two data points is the relative difference of the gene in its regulatory relationships in the two scGRNs. Ranked genes are subject to tests for their significance in differential regulation between scGRNs. (**e**) The quantile-quantile (q-q) plot for genes sorted by the significance in the differential regulation test of scTenifoldNet. The expected and observed fold-changes (FCs) between the distance of each gene's two projections on the manifold and the average distance are shown. Inset: the GSEA result shows that the Malaria pathway with FDR = 0.02. (**f**) The Malaria pathway in KEGG database is shown with genes differentially regulated and differentially expressed between COVID-19 patients with moderate and severe symptoms highlighted. Genes with expression is upregulated (or downregulated) in severely affected patients are highlighted in red (or green). Five key genes: *TLR2/4*, *TGFB*, *CXCL8* (i.e., *IL8*), *MCP1*, and *IL1A* (i.e., *IL1*), are further highlighted with an orange background in the pathway.