# Automate bioinformatics analysis for SARS-CoV-19 report

## Introduction

The biological mechanism and genomic information are of extreme importance to identify therapeutic options and optimize treatment. Given a disease name, a bioinformatic research may return hundreds of results which can be challenging to summarize, not to mention keep pace with rapidly expanding biological datasets. With appropriate bioinformatic tools and resource, we sought to create an analysis pipeline to extract relevant topics and select those for key component identification and genome differences/variance analysis.

There have been 161 million confirmed cases of COVID-19 and 3,352,109 deaths (3.78% mortality rate), spread 216 countries, as of 12 May 2021[1]. The magnitude of the pandemic underscores the urgency for better prophylaxis, diagnosis, and treatment. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes COVID-19, is a member of the coronaviridae family, seven of which a known to infect humans. Four species viz.229E, OC43, NL63, and HKU1 have been found to be mildly pathogenic, while the other two species – SARS-CoV and Middle East Respiratory Syndrome (MERS)-CoV spanned over 32 and 27 countries during the period of November 2002 to August 2003 and April 2012 to December 2019 with 8422 and 2496 confirmed cases and 10.87% and 34.77% mortality rate, respectively[2–4]. Epidemiologically, COVID-19 is highly infectious to all age groups with symptoms vary greatly among individuals, causes fatal illness, including Acute respiratory distress syndrome (ARDS), encephalitis, neurologic disease, myocarditis, vasculitis, and thrombosis. Patients who require ICU admission might have the cytokine storm. A deeper knowledge of the involved signal pathways may shed light on the development of new therapeutic strategies.

SARS-CoV-2 have a large single-strand RNA genome with 27-32 kilobases (kb). While the ongoing sequencing data and antigenic typing are rapidly emerging, it is crucial to understand viruses' sequence variantion and the dynamics of the pandemic.

## Material and Methods

Our pipeline is python-based platform, designed to extract information from KEGG and NCBI websites. By inputting disease name, we are able to extract disease relevant information, including the pathway map, the gene identification numbers, network IDs, and literature references from KEGG website (Figure 1). In the meantime, in combining with blast+ built in local computer, the platform could extract SARS-CoV-2 sequence from NCBI website and analyze sequence variation against the reference database built in local machine (figure 2).

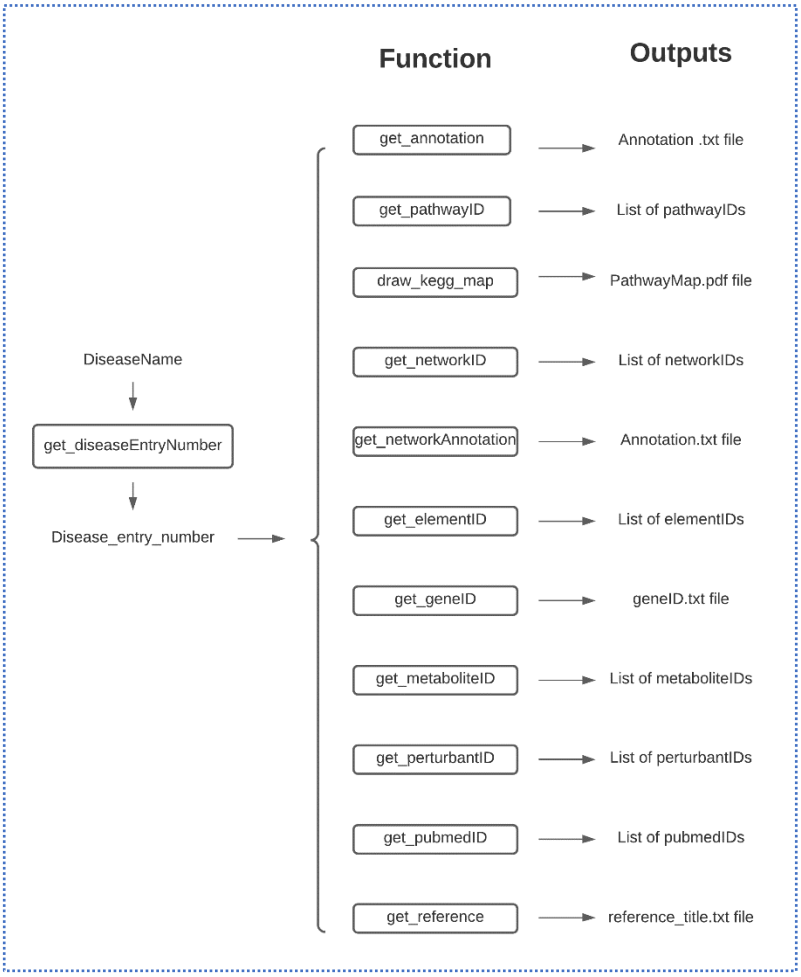



*Figure 1. flow chart for KEGG information extract*

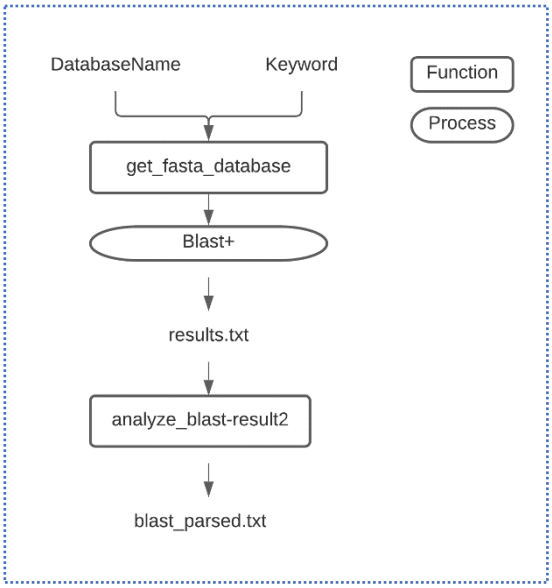



*Figure 2. flow chart for fasta file extraction and blast analysis*

**Summary**

Using our platform, we were able to extract relevant information for COVID-19. List of pathway ID, the network ID, element ID, metabolite ID, perturbant ID, and PUBMED ID were printed on the screen. Disease annotation, pathway map, network annotation, list of gene ID, and title of the references were saved in the corresponding folder.

From the aforementioned extractions, we learned the reliable, up-to-date research progress of COVID-19. The outbreak of COVID-19 introduced the highly pathogenic coronavirus into humans[5–7]. Like SARS-CoV and MERS-CoV, SARS-CoV-2 target the lower respiratory system, causing respiratory illnesses, including severe pneumonia, acute lung injury and acute respiratory distress syndrome. Infect mostly human type I and type II pneumocytes and alveolar macrophages, nasal, corneal and intestinal epithelial cells. SARS-CoV-2 uses its receptor binding domain (RBD) in spike protein subunit S1 to interact with human angiotensin converting enzyme 2 (ACE2) receptor for entry[8–10] (Figure 3). Via both structure and non-structure proteins, SARS-CoV-2 could antagonize IFN-1 production in infected monocytes and macrophages, combining with other strategies including low cytosine-phosphate(CpG) levels in the genome, glycosylation to shield the receptor-binding domain, and RNA shielding, SARS-CoV-2 is able to evade human innate immune surveillance[11–16].

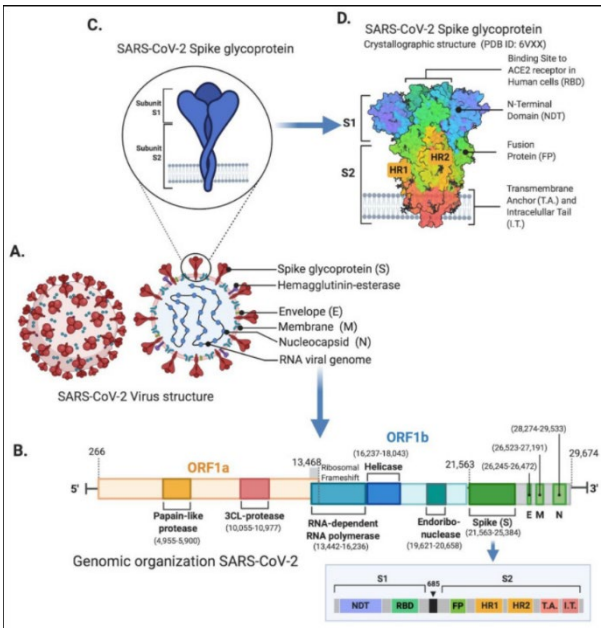



*Figure 3. Structure and genomic organization of SARS-CoV-2. (from: Chilamakuri, R., & Agarwal, S. (2021). COVID-19: Characteristics and Therapeutics. Cells, 10(2), 206.)*

After the initial phase of viral infection, 30% of hospitalized patients COVID-19 develop severe disease with progressive lung damage owing to an overreacting inflammatory response and the subsequent immuno-paralysis[17–21]. Mechanistically, several factors could induce inflammation. **First,** the virus-mediated down-regulation of ACE2 causes a burst of inflammatory cytokine release through dysregulation of the renin-angiotensin-aldosterone system, attenuation of Mas receptor, increased activation of bradykinin, and activation of the complement system[22–25]. **Second,** dysregulated innate anti-viral response may contribute to systemic hyper inflammation. Infected monocytes/macrophages by SARS-CoV-2 could impair the adaptive immune response against the virus, could migrate and spread viruses and produce large amounts of numerous types of pro-inflammatory cytokines such as IL1, IL6, IL2, IL7, IL10, G-CSF, IP10, MCP1, MIP1α, and TNFα, contribute to ARDS and cytokine storm[26–28]. Elevated systemic IL6 levels in patients strongly predict the need for intensive care[29]. **Third,** increased amount of activated DC and neutrophils could also contribute to viral spread, excessive inflammation and activation-induced lymphocytic cell death and thus decreased T cells. These accumulated innate immune disturbances not only exacerbate tissue damage in the respiratory system, but also lead to widespread activation of coagulation pathways leading to thrombosis[15,30]. **Last,** immunodeficiency involving

impaired type-1 interferon response and lymphopenia, IL1β-driven macrophage activation syndrome, and IL6-driven immune dysregulation[31].

In addition, sustained, systemic activation of the complement pathways caused significant deposits of terminal complement components was found colocalized with catastrophic microvascular injury associated with the procoagulant state in the lung tissue of COVID-19 severe patients. Unrestricted complement system activation is posited play a vital role in acute and chronic inflammation, endothelial cell dysfunction, thrombus formation and intravascular coagulation, and ultimately contributes to multiple organ failure and death[32–39].

In the end, endothelial injury caused thrombotic complication ranging from microvascular thrombosis, venous thromboembolic disease, and stroke, which are prominent manifestations of severe COVID-19. The inflammatory environment and hypercytokinemia could cause endotheliitis and thrombotic events and intravascular coagulation. The vascular barrier breach could lead to tissue edema, dissemination of the virus across the blood-brain barrier, and deregulated inflammatory cell infiltration, may play a crucial role in determining prognosis of COVID-19 patients[17,33,40–42].

Potential treatment target is focus on SARS-CoV2 PLpro[16,43], S protein-NRP1 interaction[44], IFN and its stimulated genes[11,45], ACE2 receptor[46,47], cytokine and IL6 signaling [31,40,48–51], T cells[52], thrombosis[41], and complement system [33,38].

We extract 200 S protein sequence and compare it to the SARS-CoV-2 genome sequence database built in our local machine using blast+ (Figure 4). Exceedingly high similarities with 98~100% sequence identity was found between S protein sequences and control database (Figure 5), which suggests a common source for these infections.

Use 200 fasta full genomic sequences of COVID19 to create blastdb

```
wen_x@DESKTOP-P6732JK ~
$ pwd
/home/wen_x

wen_x@DESKTOP-P6732JK ~
$ makeblastdb -in SARS-CoV-2_nucleotide_fasta.txt -dbtype nucl -parse_seqids -title "CoVID19_200Seqs" -out
 CoVID19_200


Building a new DB, current time: 05/10/2021 16:54:14
New DB name:   C:\cygwin64\home\wen_x\CoVID19_200
New DB title:  CoVID19_200Seqs
Sequence type: Nucleotide
Keep MBits: T
Maximum file size: 10000000000B
Adding sequences from FASTA; added 200 sequences in 0.0808502 seconds.

wen_x@DESKTOP-P6732JK ~
$ ls
 CoVID19_200.nhr   NC_045512.fasta                          nt.nhr
 CoVID19_200.nin  'SARS-CoV-2 S protein_nucleotide_fasta.txt'  nt.nin
 CoVID19_200.nog   SARS-CoV-2_nucleotide_fasta.txt           nt.nsq
 CoVID19_200.nsd   edirect                                   xtract.CYGWIN_NT.gz
 CoVID19_200.nsi   install-edirect.sh
 CoVID19_200.nsq   nt.fasta
```

Run a blast and look at sequence homology between the cDNA sequence of the S protein and the blastdb pairwise .txt file:

```
wen_x@DESKTOP-P6732JK ~
$ blastn -db CoVID19_200 -query 'SARS-CoV-2 S protein_nucleotide_fasta.txt' -out results.out

wen_x@DESKTOP-P6732JK ~
$
```

*Figure 4. blast+ analysis*

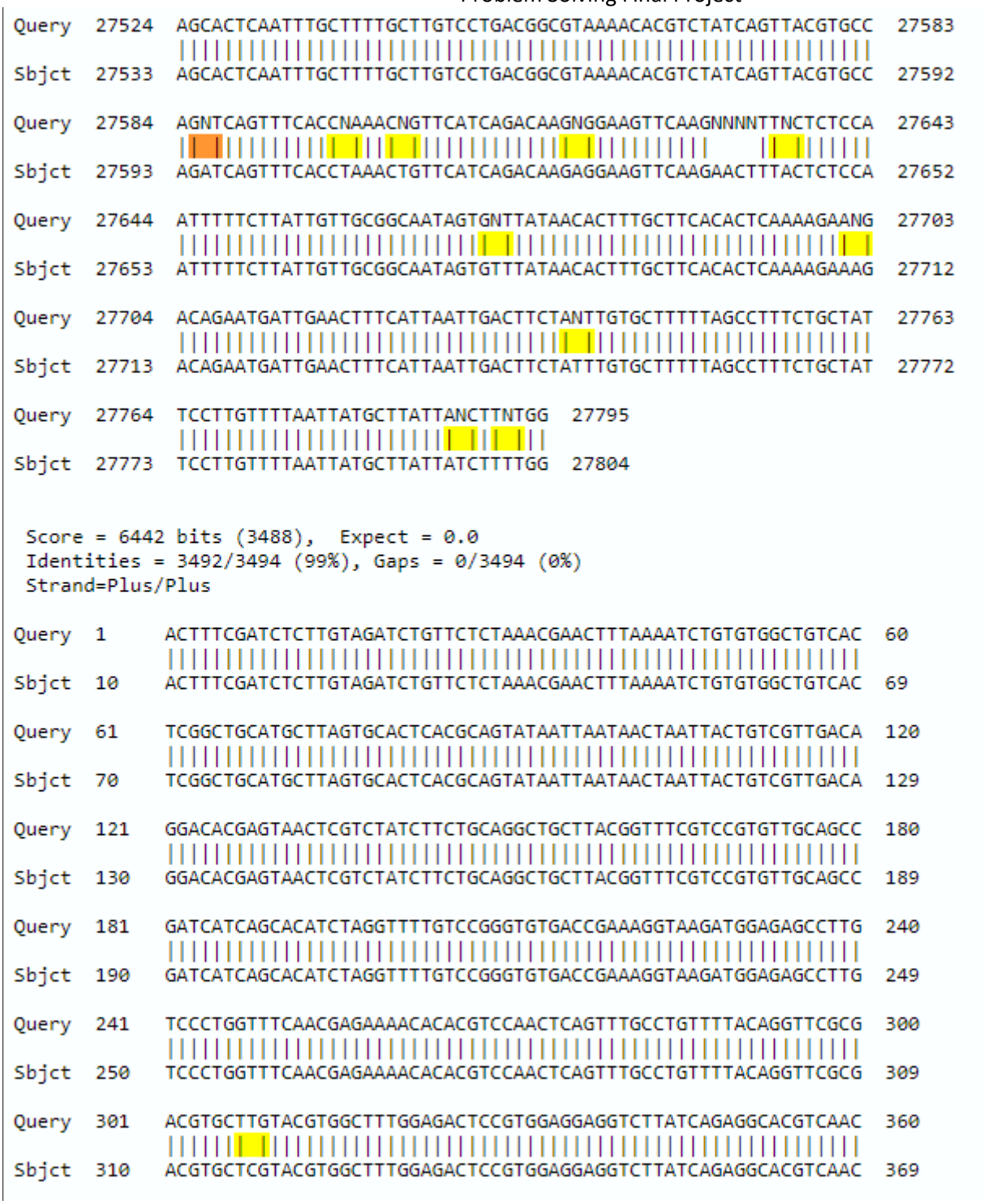
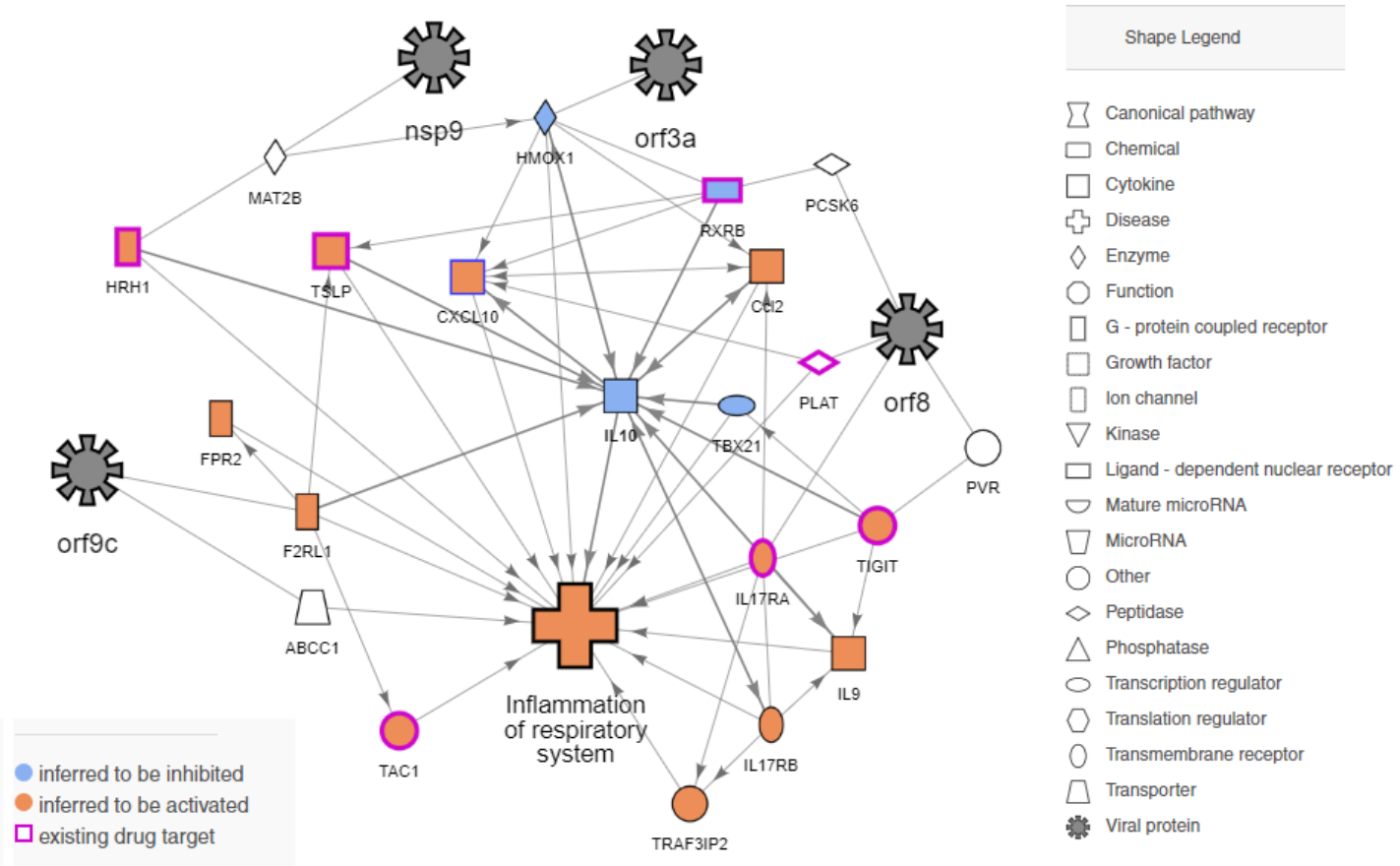



*Figure 5. screen shot of pairwise sequence alignment*

……

## Additional analysis

A molecular network built from Qiagen platform highlights possible factors that could affect inflammation of the respiratory system. Based on previously identified viral interactions with human host proteins and the machine learning model, the potential drug targets that could modulate the pulmonary inflammation process are predicted.




| Existing drug target | File name |
|---|---|
| Drugs targeting HRH1 and predicted to affect the biological process | drugs_HRH1.txt |
| Drugs targeting TSLP and predicted to affect the biological process | drug_TSLP.txt |
| Drugs targeting RXRB and predicted to affect the biological process | drugs_RXRB.txt |
| Drugs targeting PLAT and predicted to affect the biological process | drugs_PLAT.txt |
| Drugs targeting TIGIT and predicted to affect the biological process | drugs_TIGIT.txt |
| Drugs targeting IL17RA and predicted to affect the biological process | drugs_IL17RA.txt |
| Drugs targeting TAC1 and predicted to affect the biological process | drugs_TAC1.txt |

## Reference


1. WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data. https://covid19.who.int/.
2. Li, E. *et al.* Characterization of the Immune Response of MERS-CoV Vaccine Candidates Derived from Two Different Vectors in Mice. *Viruses* **12**, (2020).
3. Cui, J., Li, F. & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology* vol. 17 181–192 (2019).
4. Fehr, A. R. & Perlman, S. Coronaviruses: An overview of their replication and pathogenesis. in *Coronaviruses: Methods and Protocols* vol. 1282 1–23 (Springer New York, 2015).
5. Quan, C., Li, C., Ma, H., Li, Y. & Zhang, H. Immunopathogenesis of Coronavirus-Induced Acute Respiratory Distress Syndrome (ARDS): Potential Infection-Associated Hemophagocytic Lymphohistiocytosis. *Clin. Microbiol. Rev.* **34**, (2020).
6. Hui, D. S. *et al.* The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int. J. Infect. Dis.* **91**, 264–266 (2020).
7. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
8. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020).
9. Chilamakuri, R. & Agarwal, S. COVID-19: Characteristics and Therapeutics. *Cells* vol. 10 (2021).
10. Hartenian, E. *et al.* The molecular virology of coronaviruses. *J. Biol. Chem.* **295**, 12910–12934 (2020).
11. Lei, X. *et al.* Activation and evasion of type I interferon responses by SARS-CoV-2. *Nat. Commun.* **11**, 3810 (2020).
12. Xia, H. *et al.* Evasion of Type I Interferon by SARS-CoV-2. *Cell Rep.* **33**, 108234 (2020).
13. Harrison, A. G., Lin, T. & Wang, P. Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends Immunol.* **41**, 1100–1115 (2020).
14. Shi, M. *et al.* SARS-CoV-2 Nsp1 suppresses host but not viral translation through a bipartite mechanism. *bioRxiv Prepr. Serv. Biol.* (2020) doi:10.1101/2020.09.18.302901.
15. Amor, S., Fernández Blanco, L. & Baker, D. Innate immunity during SARS-CoV-2: evasion strategies and activation trigger hypoxia and vascular damage. *Clin. Exp. Immunol.* **202**, 193–209 (2020).
16. McClain, C. B. & Vabret, N. SARS-CoV-2: the many pros of targeting PLpro. *Signal Transduct. Target. Ther.* **5**, 223 (2020).
17. Teuwen, L.-A., Geldhof, V., Pasut, A. & Carmeliet, P. COVID-19: the vasculature unleashed. *Nat. Rev. Immunol.* **20**, 389–391 (2020).
18. Kiselevskiy, M. *et al.* Immune Pathogenesis of COVID-19 Intoxication: Storm or Silence? *Pharmaceuticals (Basel).* **13**, (2020).
19. Tay, M. Z., Poh, C. M., Rénia, L., MacAry, P. A. & Ng, L. F. P. The trinity of COVID-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.* **20**, 363–374 (2020).
20. Perico, L. *et al.* Immunity, endothelial injury and complement-induced coagulopathy in COVID-19. *Nat. Rev. Nephrol.* **17**, 46–64 (2021).
21. Catanzaro, M. *et al.* Immune response in COVID-19: addressing a pharmacological challenge by targeting pathways triggered by SARS-CoV-2. *Signal Transduct. Target. Ther.* **5**, 84 (2020).
22. Mahmudpour, M., Roozbeh, J., Keshavarz, M., Farrokhi, S. & Nabipour, I. COVID-19 cytokine storm: The anger of inflammation. *Cytokine* **133**, 155151 (2020).
23. Banu, N., Panikar, S. S., Leal, L. R. & Leal, A. R. Protective role of ACE2 and its downregulation in SARS-CoV-2 infection leading to Macrophage Activation Syndrome: Therapeutic implications. *Life Sci.* **256**, 117905 (2020).

24. Curran, C. S., Rivera, D. R. & Kopp, J. B. COVID-19 Usurps Host Regulatory Networks. *Front. Pharmacol.* **11**, 1278 (2020).

25. Furuhashi, M., Moniwa, N., Takizawa, H., Ura, N. & Shimamoto, K. Potential differential effects of renin-angiotensin system inhibitors on SARS-CoV-2 infection and lung injury in COVID-19. *Hypertens. Res.* **43**, 837–840 (2020).

26. Jafarzadeh, A., Chauhan, P., Saha, B., Jafarzadeh, S. & Nemati, M. Contribution of monocytes and macrophages to the local tissue inflammation and cytokine storm in COVID-19: Lessons from SARS and MERS, and potential therapeutic interventions. *Life Sci.* **257**, 118102 (2020).

27. Felsenstein, S., Herbert, J. A., McNamara, P. S. & Hedrich, C. M. COVID-19: Immunology and treatment options. *Clin. Immunol.* **215**, 108448 (2020).

28. Merad, M. & Martin, J. C. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. *Nat. Rev. Immunol.* **20**, 355–362 (2020).

29. Gubernatorova, E. O., Gorshkova, E. A., Polinova, A. I. & Drutskaya, M. S. IL-6: Relevance for immunopathology of SARS-CoV-2. *Cytokine Growth Factor Rev.* **53**, 13–24 (2020).

30. Ahmed-Hassan, H. *et al.* Innate Immune Responses to Highly Pathogenic Coronaviruses and Other Significant Respiratory Viral Infections. *Front. Immunol.* **11**, 1979 (2020).

31. Jamilloux, Y. *et al.* Should we stimulate or suppress immune responses in COVID-19? Cytokine and anti-cytokine interventions. *Autoimmun. Rev.* **19**, 102567 (2020).

32. Magro, C. *et al.* Complement associated microvascular injury and thrombosis in the pathogenesis of severe COVID-19 infection: A report of five cases. *Transl. Res.* **220**, 1–13 (2020).

33. Noris, M., Benigni, A. & Remuzzi, G. The case of complement activation in COVID-19 multiorgan impact. *Kidney Int.* **98**, 314–322 (2020).

34. Jodele, S. & Köhl, J. Tackling COVID-19 infection through complement-targeted immunotherapy. *Br. J. Pharmacol.* (2020) doi:10.1111/bph.15187.

35. Java, A. *et al.* The complement system in COVID-19: friend and foe? *JCI insight* **5**, e140711 (2020).

36. Lo, M. W., Kemper, C. & Woodruff, T. M. COVID-19: Complement, Coagulation, and Collateral Damage. *J. Immunol.* **205**, 1488–1495 (2020).

37. Polycarpou, A. *et al.* Rationale for targeting complement in COVID-19. *EMBO Mol. Med.* **12**, e12642 (2020).

38. Chauhan, A. J., Wiffen, L. J. & Brown, T. P. COVID-19: A collision of complement, coagulation and inflammatory pathways. *J. Thromb. Haemost.* **18**, 2110–2117 (2020).

39. Risitano, A. M. *et al.* Complement as a target in COVID-19? *Nat. Rev. Immunol.* **20**, 343–344 (2020).

40. Hojyo, S. *et al.* How COVID-19 induces cytokine storm with high mortality. *Inflamm. Regen.* **40**, 37 (2020).

41. McFadyen, J. D., Stevens, H. & Peter, K. The Emerging Threat of (Micro)Thrombosis in COVID-19 and Its Therapeutic Implications. *Circ. Res.* **127**, 571–587 (2020).

42. Colling, M. E. & Kanthi, Y. COVID-19-associated coagulopathy: An exploration of mechanisms. *Vasc. Med.* **25**, 471–478 (2020).

43. Klemm, T. *et al.* Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J.* **39**, e106275 (2020).

44. Kielian, M. Enhancing host cell infection by SARS-CoV-2. *Science* **370**, 765–766 (2020).

45. Liu, G. *et al.* ISG15-dependent Activation of the RNA Sensor MDA5 and its Antagonism by the SARS-CoV-2 papain-like protease. *bioRxiv Prepr. Serv. Biol.* (2020) doi:10.1101/2020.10.26.356048.

46. Zhang, H., Penninger, J. M., Li, Y., Zhong, N. & Slutsky, A. S. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Med.* **46**, 586–590 (2020).

47. South, A. M., Tomlinson, L., Edmonston, D., Hiremath, S. & Sparks, M. A. Controversies of renin-angiotensin system inhibition

during the COVID-19 pandemic. *Nat. Rev. Nephrol.* **16**, 305–307 (2020).

48. P aniri, A. & Akhavan-Niaki, H. Emerging role of IL-6 and NLRP3 inflammasome as potential therapeutic targets to combat COVID-19: Role of lncRNAs in cytokine storm modulation. *Life Sci.* **257**, 118114 (2020).

49. Zhang, C., Wu, Z., Li, J.-W., Zhao, H. & Wang, G.-Q. Cytokine release syndrome in severe COVID-19: interleukin-6 receptor antagonist tocilizumab may be the key to reduce mortality. *Int. J. Antimicrob. Agents* **55**, 105954 (2020).

50. Zhang, S., Li, L., Shen, A., Chen, Y. & Qi, Z. Rational Use of Tocilizumab in the Treatment of Novel Coronavirus Pneumonia. *Clin. Drug Investig.* **40**, 511–518 (2020).

51. Farooqi, F. *et al.* Treatment of Severe COVID-19 with Tocilizumab Mitigates Cytokine Storm and Averts Mechanical Ventilation During Acute Respiratory Distress: A Case Report and Literature Review. *Trop. Med. Infect. Dis.* **5**, (2020).

52. Omarjee, L. *et al.* Targeting T-cell senescence and cytokine storm with rapamycin to prevent severe progression in COVID-19. *Clin. Immunol.* **216**, 108464 (2020).