MEDWIN PUBLISHERS
Committed to Create Value for Researchers

Bioinformatics & Proteomics Open Access Journal
ISSN: 2642-6129

"Identification of SARS-CoV-2 in Human Genome based on Protein Dynamics Conversion and Target Genes Marking via Bioinformatics Approaches"

Vinita Srivastava[1], Shashi Prabha Agrawal[2], Nandita Trivedi[5], Alka Dubey[5]* , Sunil Kumar Mishra[3], Kanchan Chaurasia[4] and Swinderjeet Singh Kalra[1]

[1]Associate Professor from Department of Chemistry, D.A-V College, India
[2]Assistant Professor from Department of Zoology, D.A-V College, India
[3]Assistant Professor from Department of Chemistry, D.A-V College, India
[4]Reseach Associate Department of Chemistry, D.A-V College, India
[5]Visiting Research Associate in Bioinformatics Infrastructure Facility, India

Research Article
Volume 5 Issue 1
Received Date: November 08, 2021
Published Date: December 03, 2021
DOI: 10.23880/bpoj-16000147

***Corresponding author:** Alka Dubey, Visiting Research Associate, in Bioinformatics Infrastructure Facility (Funded by Dept. of Biotechnology, Govt. of India) India, Email: alkabioinfo964@gmail.com

Abstract

In this project our team worked on identification of target genes in human from Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and Human-to-human transmission of SARS-CoV-2 was confirmed on 20 January 2020 during the COVID-19 pandemic Transmission was initially assumed to occur primarily via respiratory droplets from coughs and sneezes within a range of about 1.8 meters (6ft). Laser light scattering experiments suggest that speaking is an additional mode of transmission and a far- reaching and under-researched one, indoors, with little air flow. Other studies have suggested that the virus may be airborne as well, with aerosols potentially being able to transmit the virus the first known infections from SARS-CoV-2 were discovered in Wuhan, China. The original source of viral transmission to humans remains unclear, as does whether the virus became pathogenic before or after the spillover event and we also found numbers of genes specific targets in human via bioinformatics approaches based on symptoms and its analysis. Identification of genes based on miRNA analysis of corona virus sequences. First we prepared some precursors sequences including analysis of genomics values after based on alignment similarly of short lengths nucleotide sequences in a form of miRNAs via bioinformatics approaches then we predicted targeted genes in a whole human genomes from the miRNA of corona virus and we identified some potential genes based on symptoms of corona virus infected peoples.

Keywords: Coronavirus; miRNAs; SARS-CoV-2; Precursors; Transmission

Abbreviations: WHO: World Health Organization; RCT: Replication-Transcription Complex; ORFs: Open Reading Frames; Mpro: Main Protease; nsps: Non-Structural Proteins; CRS: Cytokine Release Syndrome: ACE2: Angiotensin Converting Enzyme 2; ICU: Intensive Care Unit; MODS: Multiple Organ Dysfunction Syndromes; ZBD: Zinc-Binding Domain.

Introduction

At the end of 2019, a series of pneumonia cases of unknown cause emerged in Wuhan (Hubei, China) [1]. A few weeks later, in January 2020, deep sequencing analysis from lower respiratory tract samples identified a novel virus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as causative agent for that observed pneumonia cluster [2]. On February 11th, 2020, the World Health Organization (WHO) Director-General, Dr. Tedros Adhanom Ghebreyesus, named the disease caused by the SARS-CoV-2 as "COVID-19", and by March 11th, 2020 when the number of countries involved was 114, with more than 118,000 cases and over 4000 deaths, the WHO declared the pandemic status [3]. Corona Virus Disease 2019 (COVID-19) is an RNA virus, with a typical crown-like appearance under an electron microscope due to the presence of glycoprotein spikes on its envelope [4]. It is not the first time that a coronavirus causing an epidemic has been a significant global health threat: in November 2019, an outbreak of coronaviruses (CoVs) with severe acute respiratory syndrome (SARS)-CoV started in the Chinese province of Guangdong and again, in September 2012 the Middle East respiratory syndrome (MERS)-Co V appeared [5]. There are four genera of CoVs:

- α-coronavirus (alpha CoV),
- β-coronavirus (beta CoV) probably present in bats and rodents, while
- δ-coronavirus (delta CoV), and
- γ-coronavirus (gamma CoV) probably represent avian species [4-6]. Natural selection in an animal host before zoonotic transfer; and
- Natural selection in humans following zoonotic transfer [5,6]. Clinical features and risk factors are highly variable, making the clinical severity range from asymptomatic to fatal [7]. Understanding of COVID-19 is on-going.

This review aims to summarize early findings on the epidemiology, clinical features, diagnosis, management, and prevention of COVID-19.

- Epidemiology The COVID-19 epidemic expanded in early December from Wuhan, China's 7th most populous city, throughout China and was then exported to a growing number of countries. The first confirmed case of COVID-19 outside China was diagnosed on 13th January 2020 in Bangkok (Thailand) [8]. On the 2nd of March 2020, 67 territories outside mainland China had reported 8565 confirmed cases of COVID-19 with 132 deaths, as well as significant community transmission occurring in several countries worldwide, including Iran and Italy and it was declared a global pandemic by the WHO on the 11th of March 2020 [9]. The number of confirmed cases is constantly increasing worldwide and after Asian and European regions, a steep increase in cases is currently (31 March 2020) being observed in low-income countries [10]. It is problematic to quantify the exact size of this pandemia as it would necessary to count all cases including not only severe and symptomatic cases but also mild ones [11]. Unfortunately, to date, there is not a global and standard response to the pandemia and each country is facing the crisis based on their own possibilities, expertise and hypotheses. Thus, there are different criteria for testing, hospitalisation and estimating of cases making it difficult to calculate the number of people affected by epidemic. Based on the data we have so far, the estimated case fatality ratio among medically attended patients is approximately 2%, but, also in this case, a true ratio may not be known for some time [12]. Today, 31st of March 2020, based on the WHO reports, we have globally 693,224 confirmed cases and 33,106 deaths, distributed as follows: Western Pacific Region 103,775 cases and 3649 deaths, European Region 392,757 cases and 29,962 deaths, South East Asia Region 4084 cases and 158 deaths, Eastern Mediterranean Region 46,329 cases and 2813 deaths, Region of the Americas 142,081 cases and 2457 deaths and in the Africa region 3486 cases and 60 deaths [13].
- Pathophysiology and Clinical Manifestation to address the pathogenetic mechanisms of SARS-CoV-2, its viral structure and genome must be considered. Coronaviruses are enveloped positive strand RNA viruses with the largest known RNA genomes -30-32kb with a 50 -cap structure and 30 -poly-A tails. Starting from the viral RNA, the synthesis of polyprotein 1a/1ab (pp1a/pp1ab) in the host is realized [14]. The transcription works through the replication-transcription complex (RCT) organized in double-membrane vesicles and via the synthesis of subgenomic RNAs (sgRNAs) sequences of note, transcription termination occurs at transcription regulatory sequences, located between the so-called open reading frames (ORFs) that work as templates for the production of subgenomic mRNAs [15]. In the atypical CoV genome, at least six ORFs can be present. Among these, a frameshift between ORF1a and ORF1b guides the production of both pp1a and pp1ab polypeptides that are processed by virally encoded chymotrypsin-like protease (3CLpro) or main protease (Mpro), as well as one or two papain-like proteases for producing 16 non-structural proteins (nsps) [15]. Apart from ORF1a and ORF1b, other ORFs encode for structural proteins, including spike, membrane, envelope, and nucleocapsid proteins and accessory proteic chains [14,15]. Different CoVs present special structural and accessory proteins translated by dedicated sgRNAs. Pathophysiology and virulence mechanisms of CoVs, and therefore also of SARS-CoV-2 have links to the function of the nsps and structural proteins. For instance, research has underlined that nsps are able to block the host innate immune response [16]. Among the functions of the

structural proteins, the envelope has a crucial role in virus pathogenicity as it promotes viral assembly and release. Int. J. Environ. Res. Public Health 2020, 17, 2690 3 of 11. The pathogenic mechanism that produces pneumonia seems to be particularly complex [14–16]. The data so far available seem to indicate that the viral infection is capable of producing an excessive immune reaction in the host. In some cases, a reaction takes place, which as a whole is labelled a "cytokine storm". The effect is extensive tissue damage. The protagonist of this storm is interleukin 6 (IL-6). IL-6 is produced by activated leukocytes and acts on a large number of cells and tissues [17]. It is able to promote the differentiation of B lymphocytes, promotes the growth of some categories of cells, and inhibits the growth of others. It also stimulates the production of acute phase proteins and plays an important role in thermoregulation, in bone maintenance and in the functionality of the central nervous system [18]. Although the main role played by IL-6 is pro-inflammatory, it can also have anti-inflammatory effects. In turn, IL-6 increases during inflammatory diseases, infections, autoimmune disorders, cardiovascular diseases and some types of cancer [19]. It is also implicated into the pathogenesis of the cytokine release syndrome (CRS) that is an acute systemic inflammatory syndrome characterized by fever and multiple organ dysfunctions [20]. The virus might pass through the mucous membranes, especially nasal and larynx mucosa then enters the lungs through the respiratory tract. Then the virus would attack the targeting organs that express angiotensin converting enzyme 2 (ACE2), such as the lungs, heart, renal system and gastrointestinal tract [18–20]. The virus begins a second attack, causing the patient's condition to aggravate around 7 to 14 days after onset. B lymphocyte reduction may occur early in the disease, which may affect antibody production in the patient. Besides, the inflammatory factors associated with diseases mainly containing IL-6 were significantly increased, which also contributed to the aggravation of the disease around 2 to 10 days after onset. The clinical spectrum of COVID-19 varies from asymptomatic or paucisymptomatic forms to clinical conditions characterized by severe respiratory failure that necessitates mechanical ventilation and support in an intensive care unit (ICU), to multiorgan and systemic manifestations in terms of sepsis, septic shock, and multiple organ dysfunction syndromes (MODS) [21]. Asymptomatic infections have also been described, but their frequency is unknown. The main symptoms are reported in (Table 1). Pneumonia appears to be the most frequent serious manifestation of infection, characterized primarily by fever, cough, dyspnea, and bilateral infiltrates on chest imaging. [22] There are no specific clinical features that can yet reliably distinguish COVID-19 from other viral respiratory infections. Other, less common symptoms have included headaches, sore throat, and rhinorrhea. In addition to respiratory symptoms, gastrointestinal symptoms (e.g., nausea and diarrhea) have also been reported, and in some patients they may be the presenting complaint. Respiratory droplet transmission is the main route and it can also be transmitted through person-to-person contacts by asymptomatic carriers [21,22].

Annotated names with precursor sequence	Physical Constants			Thermodynamic Constants Conditions: 1 M NaCl at 25°C at pH 7.				Precursor protein sequences
	Length	Mol - W	GC-C	RlnK	deltaH	deltaG	Deltas	
>ANSV_PRE01 ATTAAAGGTTTATAC CTTCCCAGGTAACAA ACCAACCAACTTTCG ATCTCTTGTAGATCT GTTCTCTAAACGAAC TTTAAAATCTGTGTG GCTGTCACTCGGCTG CATGCTTAGTGCACT CACGCAGTATAATTA ATAAC	140	42950	39 %	33.404	1146.1	201.5	3029.4	**Pro_01** IKGLYLPR. QTNQLSIS CRSVL.TNF KICVAVTR LHA.CTHA V.LI

>ANSV_PRE02 ACATTAAAAGAAATA CTTGTCACATACAAT TGTTGTGTGATATAT TTCAATAAAAAGGAC TGGTATGATTTTGTA GAAAACCCAGATATA TTACGCGTATACGCC AACTTAGGTGAACGT GTACGCCAAGCTTTG TTAAAAACAGTACAA TTCTGTGATGCCATG CGAAATGCTGGTATT GTTGGTGTACTGACA TTAGATAATCAA	207	64090.9	34 %	33.404	1677.7	291.2	4455.7	**Pro_02** TLKEILVTYNCCD DDYFNKKDWYDF VENPDILRVYANL GERVRQALLKTVQ FCDAMRNAGIVGV LTLDNQ
>ANSV_PRE03 AAAAATCTTAGGGAA TTTGTGTTTAAGAAT ATTGATGGTTATTTT AAAATATATTCTAAG CACACGCCTATTAAT TTAGTGCGTGATCTC CCTCAGGGTTTTTCG GCTTTAGAACCATTG GTAGATTTGCCAATA GGTAT	104	43285	33%	33.404	1122.2	192.1	29.82.8	**Pro_03** KNLREFVFKNIDG YFKIYSKHTPINLV RDLPQGFSALEPL VDLPIG
>ANSV_PRE04 ACTCTGAGCCAGTGC TCAAAGGAGTCAAAT TACATTACACATAAA CGAACTTATGGATTT GTTTATGAGAATCTT CACAATTGGAACTGT AACTTTGAAGCAAGG TGAAATCAAGGATGC TACTCCTTCAGATTT TGTTCGCGCTACTGC AACGATACCGATACA AGCCTCACTCCCTTTC GGATGGCTTATTGTT GGCGTTGCACTTCT	158	48667.7	42%	33.404	1313.3	236.6	3455	**Pro_04** TLSQCSKESNYITH KRTYGFVYENLHN WNCNFEAR.NQGC YSFRFCSRYCNDT DTSLTPFRMAYC WRCTS
>ANSV_PRE05 CAATTAACACCAATA GCAGTCCAGATGACC AAATTGGCTACTACC GAAGAGCTACCAGAC GAATTCGTGGTGGTG ACGGTAAAATGAAAG ATCTCAGTCCAAGAT GGTATTTCTACTACC TAGGAACTGGGCCAG AAGCT	140	43284.2	45%	33.404	1164.3	208.5	3065.6	**Pro_05** QLTPIAVQMTKLA TTEELPDEFVVVT VK.KISVQDGISTT. ELGQK

>ANSV_PRE06 AAGGAAATTTTGGGG ACCAGGAACTAATCA GACAAGGAACTGATT ACAAACATTGGCCGC AAATTGCACAATTTG CCCCCAGCGCTTCAGC GTTCTTCGGAATGTC GCGCATTGGCATGGA AGTCACACCTTCGGG AACG	140	43260.1	49%	33.404	1216.3	222.7	1388.7	**Pro_06** KEILGTRN.SDKEL ITNIGRKLHNLPP ALQRSSECRALAW KSHLRE
>ANSV_PRE07 TGAATAAGCATATTG ACGCATACAAAACAT TCCCACCAACAGAGC CTAAAAAGGACAAAA AGAAGAAGGCTGATG AAACTCAAGCCTTAC CGCAGAGACAGAAGA AACAGCAAACTGTGA CTCTTCTTCCTGCTGC AGAT	140	43194.2	42%	33.404	1148.6	207.2	3018.3	**Pro_07** ISILTHTKHSHQQS LKRTKRRRLMKL KPYRRDRRNSKL. LFFLLQ
>ANSV_PRE08 CAGACCACACAAGGC AGATGGGCTATATAA ACGTTTTCGCTTTTC CGTTTACGATATATA GTCTACTCTTGTGCA GAATGAATTCTCGTA ACTACATAGCACAAG TAGATGTAGTTAACT TTAATCTCACATAGC AATCT	140	43062.1	37%	33.404	1120.9	196.9	2963.1	**Pro_08** QTTQGRWAI.TFS LFRLRYIVYSCAE.I LVTT.HK.M.LTLIS HSN
>ANSV_PRE09 TTAATCAGTGTGTAA CATTAGGGAGGACTT GAAAGAGCCACCACA TTTTCACCGAGGCCA CGCGGAGTACGATCG AGTGTACAGTGAACA ATGCTAGGGAGAGCT GCCTATATGGAAGAG CCCTAATGTGTAAAA TTAAT	140	43466.3	45%	33.404	1167.8	210.1	7071.9	**Pro_09** LISV.H.GGLERATT FSPRPRGVRSSVQ. TMLGRAAYMEEP. CVKL
>ANSV_PRE10 TTTAGTAGTGCTATC CCCATGTGATTTTAA TAGCTTCTTAGGAGA ATGACAAAAAAAAAA AAAAAAAAAAAAAAA AAAAAAAA	83	25731	22%	33.404	652.6	104.3	1750.9	**Pro_10** FSSAIPM.F..LLRR MTKKKKKKKKKK

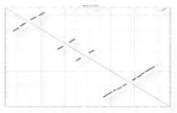
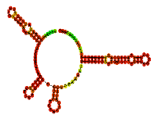
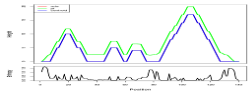
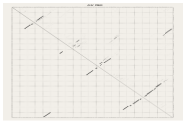
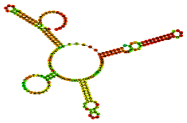
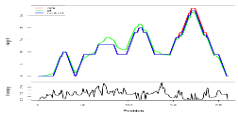
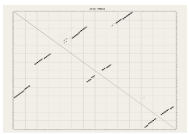
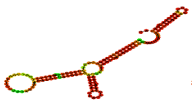
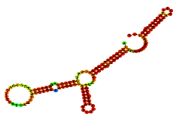
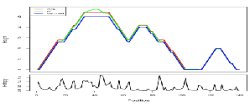
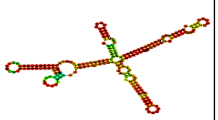
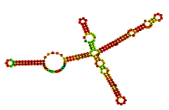
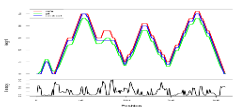
Table 1: Identified precursors with their encoded proteins including selective statically values.

Materials and Methodology

The genome sequences of corona virus were retrieved from NCBI (www.ncbi.nlm.nihgov). Precursor predicting tools, Mireval (http://tagc.univ mrs.fr/mireval) and MiRPARA (http://www.whiov.ac.cn/bioinformatics/mirpara) were used to find out the precursor sequences; and then the secondary structure with optimal minimum free energy was found out with the help of different types of web servers such as Mfold web server (http://mfold.rit.albany.edu/doc/mfold-manual/), RNA fold web server (http://rna.tbi.univie.ac.at) and DINA melt web server (http://www.microrna.gr/microT). For validating the energy values, miRBASE (mirbase@manchester.ac.uk) was checked, where we browsed all the mature virus sequences and stem loop structure sequences in FASTA format. These sequences were already experimentally identified. Then we find out the optimal minimum energy value of flu viruses in the miRBASE (only precursors), based on multiple sequence alignment, and then find out the conserved regions for miRNA. This miRNA predicted the target sites, and the target prediction was completed with the help of DIANA tar web server (http://www.microrna.gr); it predicted the different types of target sites.

Results and Discussion

This study based on bioinformatics platforms and represent. The phylogenetic analysis for the whole genome shows that SARS-CoV-2 is clustered with SARS-CoV and SARS-related coronaviruses (SARSr-CoVs) found in bats, placing it in the subgenus Sarbecovirus of the genus Beta coronavirus within this clade, SARS-CoV-2 is grouped in a distinct lineage together with four horseshoe bat coronavirus isolates (RaTG13, RmYN02, ZC45 and ZXC21) as well as novel coronaviruses recently identified in pangolins, which group parallel to SARS-CoV and other SARSr-CoVs. Using sequences of five conserved replicative domains in pp1ab (3C-like protease (3CLpro), nidovirus RNA-dependent RNA polymerase (RdRp)-associated nucleotidyltransferase (NiRAN), RdRp, zinc-binding domain (ZBD) and HEL1), the Coronaviridae Study Group of the International Committee on Taxonomy of Viruses estimated the pairwise patristic distances between SARS-CoV-2 and known coronaviruses, and assigned SARS-CoV-2 to the existing species SARSr-CoV-2. Although phylogenetically related, SARS-CoV-2 is distinct from all other coronaviruses from bats and pangolins in this species. To assess the genetic variation of different SARS-CoV-2 strains, the 2019 Novel Coronavirus Resource of China National Center for Bioinformation aligned 77,801 genome sequences of SARS-CoV-2 detected globally and identified a total of 15,018 mutations, including 14,824 single-nucleotide polymorphisms. In the S protein, four amino acid alterations, V483A, L455I, F456V and G476S, are located near the binding interface in the RBD, but their effects on binding to the host receptor are unknown. But in this study we found 20 specific gen nes they target in humabody via corona virus receptors as given below in tables (Tables 2-5).

Precursor number	MFE (kcal/mol.)	FEE (kcal/mol.)	MFE of C (kcal/mol.)	Dot plot with base pair probabilities	MFE secondary structure	Centroid SS	Mountain plot MFE structure
>ANSV_PRE01	-35.1	-36.82	-35.10				
>ANSV_PRE02	-46.80	-51.70	-46.10				
>ANSV_PRE03	-38.40	-41.01	-35.50				
>ANSV_PRE04	-51.90	-55.63	-47.90				

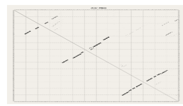
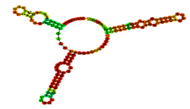
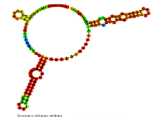
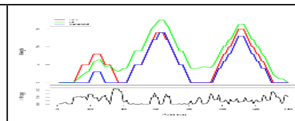
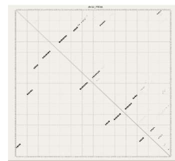
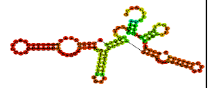
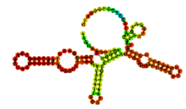
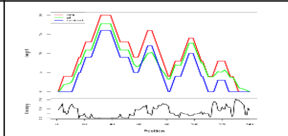
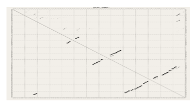
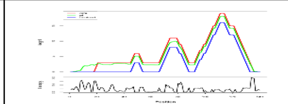
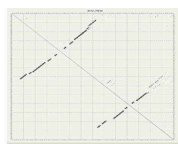
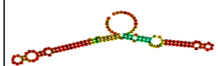
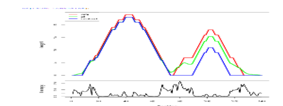
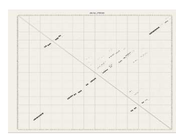
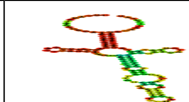
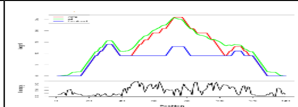
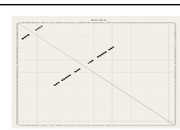
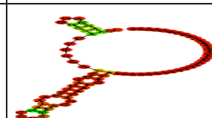
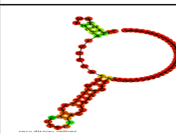
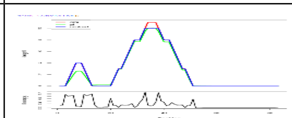
>ANSV_PRE05	-33.40	-35.67	-23.90				
>ANSV_PRE06	-35.80	-37.52	-33.70				
>ANSV_PRE07	-27.30	-29.49	-25.70				
>ANSV_PRE08	-30.40	-32.93	-29.00				
>ANSV_PRE09	-28.40	-31.35	-19.80				
>ANSV_PRE10	-5.40	-6.45	-5.30				

Table 2: Secondary structures of precursors and their plots and MFEs separately.

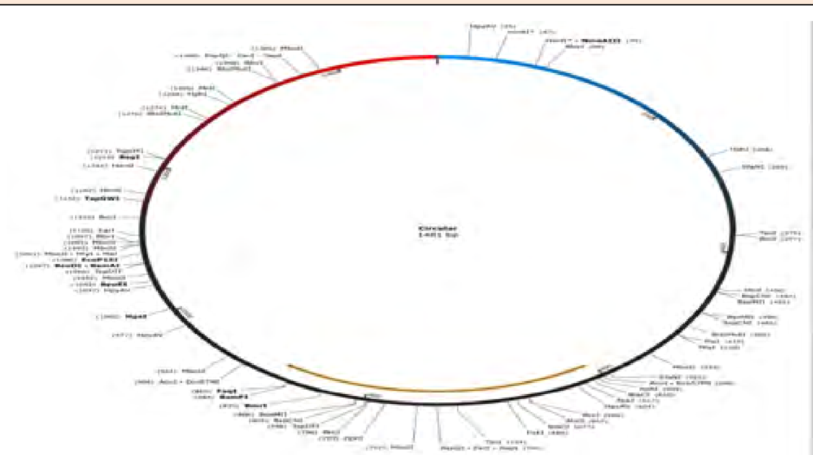
Compiled Selected Precursors Sequences	Structure of Compiled Precursors Sequences with Position of Enzymes and Genes
Pre_com _01 >ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACTTTAAAATCT GTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAATAACACATTAAAAGAAATACTTGTCACATACAATT GTTGTGATGATGATTATTTCAATAAAAAGGACTGGTATATTTTGTAGAAAACCCAGATATATTACGCGTATACGCCAACTTAGGT GAACGTGTACGCCAAGCTTTGTTAAAAACAGTACAATTCTGTGATGCCATGCGAAATGCTGGTATTGTTGGTGTACTGACATTAGA TAATCAAAAAAATCTTAGGGAATTTGTGTTTAAGAATATTGATGGTTATTTTAAAATATATTCTAAGCACACGCCTATTAATTTA GTGCGTGATCTCCCTCAGGGTTTTTCGGCTTTAGAACCATTGGTAGATTTGCCAATAGGTATACTCTGAGCCAGTGCTCAAAGGAG TCAAATTACATTACACATAAACGAACTTATGGATTTGTTTATGAGAATCTTCACAATTGGAACTGTAACTTTGAAGCAAGGTGAAA TCAAGGATGCTACTCCTTCAGATTTTGTTCGCGCTACTGCAACGATACCGATACAAGCCTCACTCCCTTTCGGATGGCTTATTGTTG GCGTTGCACTTCTCAATTAACACCAATAGCAGTCCAGATGACCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGT GACGGTAAAATGAAAGATCTCAGTCCAAGATGGTATTTCTACTACCTAGGAACTGGGCCAGAAGCAAGGAAATTTTGGGGACCAGG AACTAATCAGACAAGGAACTGATTACAAACATTGGCCGCAAATTGCACAATTTGCCCCCAGCGCTTCAGCGTTCTTCGGAATGTCGC GCATTGGCATGGAAGTCACACCTTCGGGAACGTGAATAAGCATATTGACGCATACAAAACATTCCCACCAACAGAGCCTAAAAAGG ACAAAAAGAAGAAGGCTGATGAAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCAAACTGTGACTCTTCTTCCTGCTGCAGAT CAGACCACACAAGGCAGATGGGCTATATAAACGTTTTCGCTTTTCCGTTTACGATATATAGTCTACTCTTGTGCAGAATGAATTCT CGTAACTACATAGCACAAGTAGATGTAGTTAACTTTAATCTCACATAGCAATCTTTAATCAGTGTGTAACATTAGGGAGGACTTGA AAGAGCCACCACATTTTCACCGAGGCCACGCGGAGTACGATCGAGTGTACAGTGAACAATGCTAGGGAGAGCTGCCTATATGGAAG AGCCCTAATGTGTAAAATTAATTTTAGTAGTGCTATCCCCATGTGATTTTAATAGCTTCTTAGGAGAATGACAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAA	

Table 3: Compilation of the genomic and enzymatic functions of identified genes.

Compiled Sequence of Precursors Proteins via Translation	Structure of Compiled Sequence of Precursors Proteins via Translation
>CPPS_01 IKGLYLPRQTNQLSISCRSVLTNFKICVAVTRLHACTHAVLITLKEILVT YNCCDDDYFNKKDWYDFVENPDILRVYANLGERVRQALLKTVQFCD AMRNAGIVGVLTLDNQ KNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIG TLSQCSKESNYITHKRTYGFVYENLHNWNCNFEARNQGCYSFRFCSRY CNDTDTSLTPFRMAYCWRCTSQLTPIAVQMTKLATTEELPDEFVVVT VKKISVQDGISTTELGQKKEILGTRNSDKELITNIGRKLHNLPPALQRS SECRALAWKSHLRISILTHTKHSHQQSLKRTKRRRLMKLKPYRRDRR NSKLLFFLLQQTTQGRWAITFSLFRLRYIVYSCAEILVTTHKMLTLISH SNLISVHGGLERATTFSPRPRGVRSSVQTMLGRAAYMEEPCVKL FSSAIPMFLLRRMTKKKKKKKKKK	

Table 4: Structure of compiled protein sequences with amino acid position based on full genomic transformation.

S.No	Gene name	Function of Identified Genes
1	ACE2	The encoded protein is a functional receptor for the spike glycoprotein of the human coronavirus HCoV-NL63 and the human severe acute respiratory syndrome coronaviruses, SARS-CoV and SARS-CoV-2, the latter is the causative agent of coronavirus disease-2019 (COVID-19). Multiple splice variants have been found for this gene and the dACE2 (or MIRb-ACE2) splice variant has been found to be interferon inducible.
2	2 IBVgp2	Infectious bronchitis virus; spike protein.
3	TMPRSS2	This gene encodes a protein that belongs to the serine protease family. The encoded protein contains a type II transmembrane domain, a receptor class A domain, a scavenger receptor cysteine-rich domain and a protease domain. Serine proteases are known to be involved in many physiological and pathological processes. This gene was demonstrated to be up-regulated by androgenic hormones in prostate cancer cells and down-regulated in androgen-independent prostate cancer tissue. T
4	N- GU280_gp10	Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an envelope, positive-sense, single-stranded RNA virus that causes coronavirus disease 2019 (COVID-19). Virus particles include the RNA genetic material and structural proteins needed for invasion of host cells. Once inside the cell the infecting RNA is used to encode structural proteins that make up virus particles, nonstructural proteins that direct virus assembly, transcription, replication and host control and accessory proteins whose function has not been determined.~ The structural proteins of SARS-CoV-2 include the envelope protein (E), spike or surface glycoprotein (S), membrane protein (M) and the nucleocapsid protein (N).
5	S -GU280_gp02	Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an enveloped, positive-sense, single-stranded RNA virus that causes coronavirus disease 2019 (COVID-19). Virus particles include the RNA genetic material and structural proteins needed for invasion of host cells. Once inside the cell the infecting RNA is used to encode structural proteins that make up virus particles, nonstructural proteins that direct virus assembly, transcription, replication and host control and accessory proteins whose function has not been determined.~ The structural proteins of SARS-CoV-2 include the envelope protein (E), spike or surface glycoprotein (S), membrane protein (M) and the nucleocapsid protein (N).

6	E -GU280_gp04	Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an enveloped, positive-sense, single-stranded RNA virus that causes coronavirus disease 2019 (COVID-19). Virus particles include the RNA genetic material and structural proteins needed for invasion of host cells.
7	ORF8- GU280_gp09	Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an enveloped, positive-sense, single-stranded RNA virus that causes coronavirus disease 2019 (COVID-19). Virus particles include the RNA genetic material and structural proteins needed for invasion of host cells. Once inside the cell the infecting RNA is used to encode structural proteins that make up virus particles, nonstructural proteins that direct virus assembly, transcription, replication and host control and accessory proteins whose function has not been determined.~ ORF8 encodes a viral accessory protein.
8	ORF7b-GU280_gp08	Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an enveloped, positive-sense, single-stranded RNA virus that causes coronavirus disease 2019 (COVID-19). Virus particles include the RNA genetic material and structural proteins needed for invasion of host cells. Once inside the cell the infecting RNA is used to encode structural proteins that make up virus particles, nonstructural proteins that direct virus assembly, transcription, replication and host control and accessory proteins whose function has not been determined.~ ORF7b encodes a viral accessory protein. Based on its similarity to other coronavirus proteins, ORF7b protein is thought to localize to the Golgi compartment.
9	HCoV229Egp1	(replicase polyprotein 1a; replicase polyprotein 1ab) contains proteinase responsible for cleavage of the polyprotein and several NSPs . This protein has a huge multienzyme complex of several cellular proteins and 16nsp (Nonstructural protein) each nsp has its own function and contribution for virus sustainability
10	TNF	This gene encodes a multifunctional proinflammatory cytokine that belongs to the tumor necrosis factor (TNF) superfamily. This cytokine is mainly secreted by macrophages. It can bind to, and thus functions through its receptors TNFRSF1A/TNFR1 and TNFRSF1B/TNFBR. This cytokine is involved in the regulation of a wide spectrum of biological processes including cell proliferation, differentiation, apoptosis, lipid metabolism, and coagulation. This cytokine has been implicated in a variety of diseases, including autoimmune diseases, insulin resistance, psoriasis, rheumatoid arthritis ankylosing spondylitis, tuberculosis, autosomal dominant polycystic kidney disease, and cancer.
11.	NRP1	This gene encodes one of two neuropilins, which contain specific protein domains which allow them to participate in several different types of signaling pathways that control cell migration. Neuropilins contain a large N-terminal extracellular domain, made up of complement-binding, coagulation factor V/VIII, and meprin domains. These proteins also contains a short membrane-spanning domain and a small cytoplasmic domain. Neuropilins bind many ligands and various types of co-receptors; they affect cell survival, migration, and attraction. Some of the ligands and co-receptors bound by neuropilins are vascular endothelial growth factor (VEGF) and semaphorin family members
12.	CXCL5	This gene encodes a protein that is a member of the CXC subfamily of chemokines. Chemokines, which recruit and activate leukocytes, are classified by function (inflammatory or homeostatic) or by structure. This protein is proposed to bind the G-protein coupled receptor chemokine (C-X-C motif) receptor 2 to recruit neutrophils, to promote angiogenesis and to remodel connective tissues. This protein is thought to play a role in cancer cell proliferation, migration, and invasion.
13.	IL6	This gene encodes a cytokine that functions in inflammation and the maturation of B cells. In addition, the encoded protein has been shown to be an endogenous pyrogen capable of inducing fever in people with autoimmune diseases or infections. The protein is primarily produced at sites of acute and chronic inflammation, where it is secreted into the serum and induces a transcriptional inflammatory response through interleukin 6 receptor, alpha.

14.	TMPRSS2	This gene encodes a protein that belongs to the serine protease family. The encoded protein contains a type II transmembrane domain, a receptor class A domain, a scavenger receptor cysteine-rich domain and a protease domain. Serine proteases are known to be involved in many physiological and pathological processes. This gene was demonstrated to be up-regulated by androgenic hormones in prostate cancer cells and down-regulated in androgen-independent prostate cancer tissue. The protease domain of this protein is thought to be cleaved and secreted into cell media after autocleavage.
15.	AGT	The protein encoded by this gene, pre-angiotensinogen or angiotensinogen precursor, is expressed in the liver and is cleaved by the enzyme renin in response to lowered blood pressure. The resulting product, angiotensin I, is then cleaved by angiotensin converting enzyme (ACE) to generate the physiologically active enzyme angiotensin II. The protein is involved in maintaining blood pressure, body fluid and electrolyte homeostasis, and in the pathogenesis of essential hypertension and preeclampsia.
16.	TLR3	The protein encoded by this gene is a member of the Toll-like receptor (TLR) family which plays a fundamental role in pathogen recognition and activation of innate immunity. TLRs are highly conserved from Drosophila to humans and share structural and functional similarities. They recognize pathogen-associated molecular patterns (PAMPs) that are expressed on infectious agents, and mediate the production of cytokines necessary for the development of effective immunity.
17.	FGF2	The protein encoded by this gene is a member of the fibroblast growth factor (FGF) family. FGF family members bind heparin and possess broad mitogenic and angiogenic activities. This protein has been implicated in diverse biological processes, such as limb and nervous system development, wound healing, and tumor growth. The mRNA for this gene contains multiple polyadenylation sites, and is alternatively translated from non-AUG (CUG) and AUG initiation codons, resulting in five different isoforms with distinct properties. The CUG-initiated isoforms are localized in the nucleus and are responsible for the intracrine effect, whereas, the AUG-initiated form is mostly cytosolic and is responsible for the paracrine and autocrine effects of this FGF.
18.	OAS1	This gene is induced by interferons and encodes a protein that synthesizes 2',5'-oligoadenylates (2-5As). This protein activates latent RNase L, which results in viral RNA degradation and the inhibition of viral replication. Alternative splicing results in multiple transcript variants with different enzymatic activities. Polymorphisms in this gene have been associated with susceptibility to viral infection and diabetes mellitus, type 1. A disease-associated allele in a splice acceptor site influences the production of the p46 splice isoform. This gene is located in a cluster of related genes on chromosome 12
19.	GU280_gp02	Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an envelope, positive-sense, single-stranded RNA virus that causes coronavirus disease 2019 (COVID-19). Virus particles include the RNA genetic material and structural proteins needed for invasion of host cells.
20.	DPP4	The DPP4 gene encodes dipeptidyl peptidase 4, which is identical to adenosine deaminase complexing protein-2, and to the T-cell activation antigen CD26. It is an intrinsic type II transmembrane glycoprotein and a serine exopeptidase that cleaves X-proline dipeptides from the N-terminus of polypeptides.

Table 5: Identified name of the genes with their functions in human body as target genes from host of coronavirus.

Conclusion

20 numbers of identified genes fully targeted in human immune system and SARS-CoV-2 uses ACE2 as the receptor and human proteases as entry activators; subsequently it fuses the viral membrane with the cell membrane and achieves invasion. Thus, drugs that interfere with entry may be a potential treatment for COVID-19. Umifenovir (Arbidol) is a drug approved in Russia and China for the treatment of influenza and other respiratory viral infections. It can target the interaction between the S protein and ACE2 and inhibit membrane fusion. In vitro experiments showed that

it has activity against SARS-CoV-2, and current clinical data revealed it may be more effective than lopinavir and ritonavir in treating COVID-19. However, other clinical studies showed umifenovir might not improve the prognosis of or accelerate SARS-CoV-2 clearance in patients with mild to moderate COVID-19. Yet some ongoing clinical trials are evaluating its efficacy for COVID-19 treatment. Camostat mesylate is approved in Japan for the treatment of pancreatitis and postoperative reflux oesophagitis. Previous studies showed that it can prevent SARS-CoV from entering cells by blocking TMPRSS2 activity and protect mice from lethal infection with SARS-CoV in a pathogenic mouse model (wild-type mice infected with a mouse-adapted SARS-CoV strain. Recently, a study revealed that camostat mesylate blocks the entry of SARS-CoV-2 into human lung cells thus, it can be a potential antiviral drug against SARS-CoV-2 infection, although so far there are not sufficient clinical data to support its efficacy. SARS-CoV-2 triggers a strong immune response which may cause cytokine storm syndrome. Thus, immunomodulatory agents that inhibit the excessive inflammatory response may be a potential adjunctive therapy for COVID-19. Dexamethasone is a corticosteroid often used in a wide range of conditions to relieve inflammation through its anti-inflammatory and immunosuppressant effects. Recently, the RECOVERY trial found dexamethasone reduced mortality by about one third in hospitalized patients with COVID-19 who received invasive mechanical ventilation and by one fifth in patients receiving oxygen. By contrast, no benefit was found in patients without respiratory suppor COVID-19 is the third highly pathogenic human coronavirus disease to date. Although less deadly than SARS and MERS, the rapid spreading of this highly contagious disease has posed the severest threat to global health in this century. The SARS-CoV-2 outbreak has lasted for more than half a year now, and it is likely that this emerging virus will establish a niche in humans and coexist with us for a long time. Before clinically approved vaccines are widely available, there is no better way to protect us from SARS-CoV-2 than personal preventive behaviors such as social distancing and wearing masks, and public health measures, including active testing, case tracing and restrictions on social gatherings. Despite a flood of SARS-CoV-2 research published every week, current knowledge of this novel coronavirus is just the tip of the iceberg. The animal origin and cross-species infection route of SARS-CoV-2 are yet to be uncovered. The molecular mechanisms of SARS-CoV-2 infection pathogenesis and virus–host interactions remain largely unclear. Intensive studies on these virological profiles of SARS-CoV-2 will provide the basis for the development of preventive and therapeutic strategies against COVID-19. Moreover, continued genomic monitoring of SARS-CoV-2 in new cases is needed worldwide, as it is important to promptly identify any mutation that may result in phenotypic changes of the virus. Finally, COVID-19 is challenging all human beings. Tackling this epidemic is a long-term job which requires efforts of every individual and international collaboration by scientists, authorities and the public [23-39] (Figure 1).

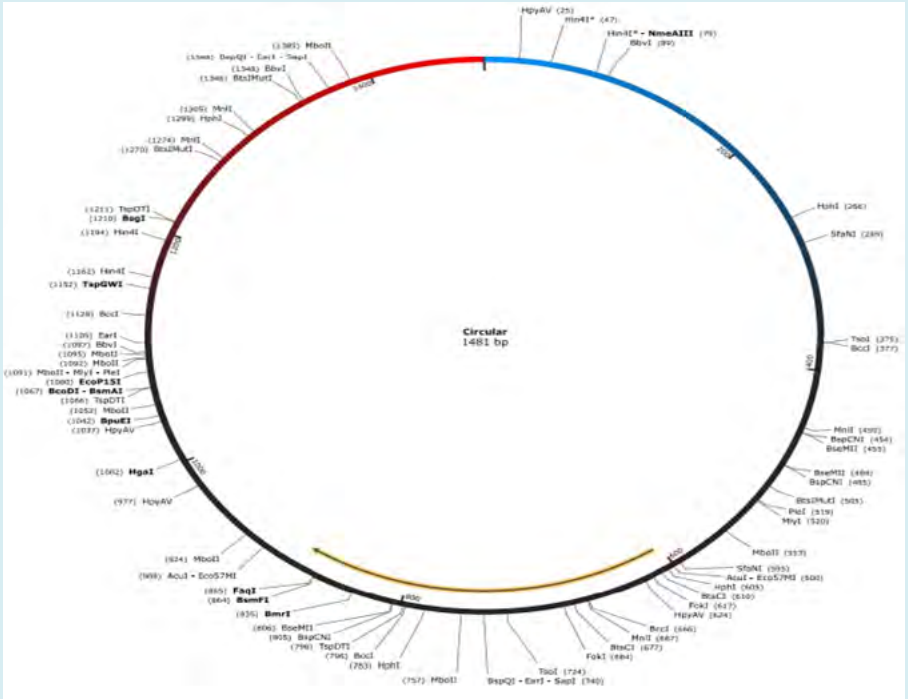


Figure 1: Position of enzymes with the name of target gene.

References

1. Lu H, Stratton CW, Tang YW (2020) Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. J Med Virol 92(4): 401-402.
2. Huang C, Wang Y, Li Z, Ren L, Zhao J, et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395(10223): 497-506.
3. (2020) World Health Organization Director-General's Opening Remarks at the Media Briefing on COVID-19-11 March 2020.
4. Perlman S, Netland J (2009) Coronaviruses post-SARS: Update on replication and pathogenesis. Nat Rev Microbiol 7(6): 439-450.
5. Lu R, Zhao X, Li J, Niu P, Yang B, et al. (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. Lancet 395(10224): 565-574.
6. Yin Y, Wunderink RG (2018) MERS, SARS and other coronaviruses as causes of pneumonia. Respirology 23(2): 130-137.
7. Phan T (2020) Novel coronavirus: From discovery to clinical diagnostics. Infect Genet Evol 79: 104211.
8. (2020) Novel Coronavirus (2019-nCoV), Situation Report-1 21 January 2020. World Health Organization.
9. Hsu LY, Chia PY, Lim JF (2020) The Novel coronavirus (SARS-CoV-2) epidemic. Ann Acad Med Singap 49(3): 105-107.
10. (2020) Coronavirus Disease 2019 Situation Report-66. World Health Organization.
11. Lipsitch M, Swerdlow DL, Finelli L (2020) Defining the epidemiology of covid-19-Studies needed. N Engl J Med 382(13): 1194-1196.
12. Lipsitch M, Donnelly CA, Fraser C, Blake IM, Cori A, et al. (2015) Potential biases in estimating absolute and relative case-fatality risks during outbreaks. PLoS Negl Trop Dis 9(7): e0003846.
13. (2020) Disease 2019 Situation Report 70. World Health Organization Coronavirus.
14. Lei J, Kusov Y, Hilgenfeld R (2018) Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. Antiviral Res 149: 58-74.
15. Letko M, Marzi A, Munster V (2020) Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol 5(4): 562-569.
16. Cascella M, Rajnik M, Aleem A, Dulebohn SC, Di Napoli R (2020) Features, Evaluation and Treatment Coronavirus (COVID-19). In StatPearls, Treasure Island, FL, USA.
17. Pyle CJ, Uwadiae FI, Swieboda DP, Harker JA (2017) Early IL-6 signalling promotes IL-27 dependent maturation of regulatory T cells in the lungs and resolution of viral immunopathology. PLoS Pathog 13(9): e1006640.
18. Chen C, Zhang XR, Ju ZY, He WF (2020) Advances in the research of cytokine storm mechanism induced by Corona Virus Disease 2019. Zhonghua Shao Shang Za Zhi 36(6): 471-475.
19. Bennardo F, Buffone C, Giudice A (2020) New therapeutic opportunities for COVID-19 patients with Tocilizumab: Possible correlation of interleukin-6 receptor inhibitors with osteonecrosis of the jaws. Oral Oncol 106: 104659.
20. John RS (2018) Interleukin-6 family cytokines. Cold Spring Harb Perspect Biol 10(2): a028415.
21. Lupia T, Scabini S, Pinna SM, Perri GD, Rosa FGD, et al. (2020) 2019 novel coronavirus (2019-nCoV) outbreak: A new challenge. J Glob Antimicrob Resist 21: 22-27.
22. Yang Y, Peng F, Wang R, Yange M, Guan K, et al. (2020) The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. J Autoimmun 109: 102434.
23. Stockman LJ, Bellamy R, Garner P (2006) SARS: systematic review of treatment effects. PLoS Med 3(9): e343.
24. Mantlo E, Bukreyeva N, Maruyama J, Paessler S, Huang C (2020) Antiviral activities of type I interferons to SARS-CoV-2 infection. Antiviral Res 179: 104811.
25. Sallard E, Lescure FX, Yazdanpanah Y, Mentre F, Smadja NP (2020) Type 1 interferons as a potential treatment against COVID-19. Antiviral Res 178: 104791.
26. Duan K, Liu B, Li C, Zhang H, Yu T, et al. (2020) Effectiveness of convalescent plasma therapy in severe COVID-19 patients. Proc Natl Acad Sci USA 117(17): 9490-9496.
27. Shen C, Wang Z, Zhao F, Yang Y, Li J, et al. (2020) Treatment of 5 critically ill patients with COVID-19 with convalescent plasma. JAMA 323(16): 1582-1589.
28. Wang C, Li W, Drabek D, Okba NMA, Haperen RV, et al.

(2020) A human monoclonal antibody blocking SARS-CoV-2 infection. Nat Commun 11(1): 2251.

29. Wu Y, Wang F, Shen C, Peng W, Li D, et al. (2020) A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. Science 368(6496): 1274-1278.

30. Zost SJ, Gilchuk P, Crowe JE (2020) Potently neutralizing and protective human antibodies against SARS-CoV-2. Nature 584: 443-449.

31. Shi R, Shan C, Duan X, Chen Z, Liu P, et al. (2020) A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. Nature 584(7819): 120-124.

32. Smith TRF, Patel A, Ramos S, Elwood D, Zhu X, et al. (2020) Immunogenicity of a DNA vaccine candidate for COVID-19. Nat Commun 11(1): 2601.

33. Zhu FC, Li YH, Guan XH, Hou LH, Wang WJ, et al. (2020) Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19 vaccine: a dose-escalation, open-label, non-randomised, first-in-human trial. Lancet 395(10240): 1845-1854.

34. Gao Q, Bao L, Mao H, Wang L, Xu K, et al. (2020) Development of an inactivated vaccine candidate for SARS-CoV-2. Science 369(6499): 77-81.

35. Zhu FC, Guan XH, Li YH, Huang JY, Jiang T, et al. (2020) Immunogenicity and safety of a recombinant adenovirus type-5-vectored COVID-19 vaccine in healthy adults aged 18 years or older: a randomised, double-blind, placebo-controlled, phase 2 trial. Lancet 396(10249): 479-488.

36. Folegatti PM, Ewer KJ, Aley PK, Angus B, Becker S, et al. (2020) Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. Lancet 396(10249): 467-478.

37. Jackson LA, Anderson EJ, Rouphael NG, Roberts PC, Makhene M, et al. (2020) An mRNA vaccine against SARSCoV-2 - preliminary report. N Engl J Med 383:1920-1931.

38. Xia S, Duan K, Zhang Y, Zhao D, Zhang H, et al. (2020) Effect of an inactivated vaccine against SARS-CoV-2 on safety and immunogenicity outcomes: interim analysis of 2 randomized clinical trials. JAMA 324(10): 951-960.

39. Tang D, Comish P, Kang R (2020) The hallmarks of COVID-19 disease. PLoS Pathog 16(5): e1008536.