Recognition of Plausible Therapeutic Agents to Combat COVID-19: An Omics Data Based Combined Approach

Mohammad Uzzal Hossain
National Institute of Biotechnology, Bioinformatics Division
Arittra Bhattacharjee
National Institute of Biotechnology, Bioinformatics Division and Department of Biochemistry and Microbiology, North South University
Md. Tabassum Hossain Emon
Department of Biotechnology and Genetic Engineering, Life Science Faculty, Mawlana Bhashani Science and Technology University
Zeshan Mahmud Chowdhury
Department of Biochemistry and Microbiology, North South University
Md. Golam Mosaib
Department of Biochemistry and Molecular Biology, Faculty of Health & Medical Sciences, Gono Bishwabidyaloy, Ashulia, Savar
Muntahi Mourin
Department of Biochemistry and Microbiology, North South University
Keshob Chandra Das
Molecular Biotechnology Division, National Institute of Biotechnology
Chaman Ara Keya
Department of Biochemistry and Microbiology, North South University
Md. Salimullah (salim2969@gmail.com)
Molecular Biotechnology Division, National Institute of Biotechnology

Research Article

Keywords: COVID-19, SARS-CoV-2, chimeric vaccine, small molecule drugs, siRNAs, Interferons

Posted Date: April 30th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-25807/v1

Version of Record: A version of this preprint was published at Gene on March 1st, 2021. See the published version at https://doi.org/10.1016/j.gene.2020.145368.

Abstract

Coronavirus disease-2019 (COVID-19) has become an immense threat to global public health. The causative agent of this disease is a novel zoonotic pathogen called Severe Acute Respiratory Syndrome related Coronavirus-2 (SARS-CoV-2). Since this is a newly evolved pathogen, very limited information is available to develop effective therapeutics against this deadly virus. Although bioinformatics based analysis could be handy to unveil drugs or vaccines against bacteria and fungus, such approaches are hardly seen for acellular viruses. However, in this study we rationally merged several powerful *in silico* techniques and proposed prospective therapeutics based on available omics data for COVID-19. Through meticulous analysis of conserved regions of 67 SARS-CoV-2 strains, spike and membrane glycoproteins were chosen to develop and propose a chimeric vaccine against this virus. siRNAs were also designed against these glycoprotein genes to silence them. Moreover, six drug compound candidates were suggested to inhibit the conserved RNA-directed RNA polymerase protein. Finally, due to the close relationship of SARS-CoV-2 and SARS-CoV, publicly available gene expression datasets of SARS-CoV were analyzed to identify 13 immunoregulatory genes that might develop interferon based therapy. Our study will quicken the researches among pharmaceuticals, researchers and clinicians to develop rapid therapeutics for controlling this notorious pandemic disease.

Authors Mohammad Uzzal Hossain, Arittra Bhattacharjee, Md. Tabassum Hossain Emon, and Zeshan Mahmud Chowdhury contributed equally to this work.

Introduction

Coronavirus Disease–2019 (COVID–19) has drastically spread throughout the globe and become a massive threat to humanity[1]. Most of the healthcare system and governing bodies from developed countries failed to stop the spreading and infection of this pandemic disease[2]. Regrettably, the quick prevention of COVID–19 is challenging because the causative agent of this disease is a novel Severe Acute Respiratory Syndrome (SARS) related Coronavirus (CoV) or SARS-CoV–2. This virus was totally unknown before the outbreak that took place in Wuhan, China; therefore, traditional development of therapeutics will not yield a rapid solution to control this pestilence [2,3].

CoVs, members of Coronaviridae family and Coronavirinae subfamily, were considered as insignificant pathogens for humans before the outbreak of SARS occurred in 2002 and 2003 from Guangdong province, China[4]. *Coronavirinae* has been divided into four categories such as *Alpha, Beta, Gamm* and *Deltacoronavirus*[4]. These pathogens were described as the “virology backwater” since very little was known about their virulence[5]. In recent decades, they have become scandalous several times due to their zoonotic characteristics. Especially, the *Betacoronavirus* group consists of the most perilous Human CoVs (HCoVs). For example, HCoV-OC43, HCoV-HKU1, SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV) are members of *Betacoronavirus*[4].

Coronavirinae has vast genetic diversity among them but they mostly reside in Bat[7]. Therefore, this mammalian bird always possesses a serious threat for CoV outbreaks, particularly in China[8]. The possible reason behind this occurrence is that the members of *Coronavirinae* frequently go under RNA recombination inside bat and often generate novel CoVs[6,7]. SARS-CoV–2, a member of *Betacoronavirus,* also originated in bats and transmitted to

humans via unidentified intermediary animals in Wuhan, China. This transmission started the pandemic COVID–19 which was instigated on December 2019[9].

SARS-CoV–2, a positive-stranded (+) RNA virus, spread via respiratory droplets of an infected person to a healthy individual through sneezing and coughing[3]. Moreover, the presence of SARS-CoV–2 is also observed in the stool; thereby contaminated water supply might promote transmission[10]. The virions enter through the upper-respiratory tract and enter inside host by engaging Angiotensin Converting Enzyme 2 (ACE2) and Transmembrane Serine Protease 2 (TMPRSS2) of the lung cells[11]. This entry triggers innate immunity as a result inflammatory cytokines e.g., Interleukin (IL) - 2, IL–10, IL–7, Granulocyte-colony stimulating factor (GCSF), monocyte chemoattractant protein 1 (MCP1), Macrophage Inflammatory Proteins (MIP) 1A, Interferon gamma-induced protein- 10 (IP–10) and tumor necrosis factor alpha (TNFα) get elevated[9]. Consequently, the patients develop indistinguishable flu like symptoms such as cough, sore throat, headache, fatigue, myalgia, breathlessness, conjunctivitis and fever. In worst cases, patients develop pneumonia leading to respiratory failure and death[12]. To handle this condition, physicians are non-specifically applying anti-malarial Chloroquine, antiviral drug e.g. Oseltamivir, Ganciclovir, Lopinavirritonavir etc. and plasma of the patients that recovered from COVID–19[9,13,14]. However, this situation demands precise therapeutics for the quick betterment of the COVID–19 patients.

A recent *in vitro* study revealed potential clinically proven protease inhibitor that can block TMPRSS2 and consequently the entry of SARS-CoV–2[11]. Similar to this study, revealing more plausible therapeutics and druggable targets will elevate the probability to achieve successful treatments against COVID–19. *In silico* approaches are the most suitable way to execute this type of research in a short time with nearly zero cost. Powerful computational methods such as subtractive proteomics, comparative genomics or virtual screening can disclose therapeutic targets and drugs against bacteria, fungi or non-communicable diseases[15,16,17]. Unfortunately, for viral infections, this type of rational approach is hardly seen. As a result, unveiling the potential targets, novel drugs, interferons and immunodominant vaccines are extremely difficult against COVID–19.

In this study, we merged several *in silico* techniques for developing treatments against COVID–19. Here, we analyzed the genome of SARS-CoV–2 and selected several conserved regions to establish small molecules, chimeric vaccine and Small interfering RNAs (siRNAs). Moreover, we carried out a comparative genome analysis of SARS-CoV–2 with SARS- CoV and MERS-CoV to look for the genetic similarity. SARS-CoV showed close genetic relation and virulence pattern with SARS-CoV–2. Thus, a transcriptomic analysis of this related virus was performed to retrieve the interferon inducing genes for implementing interferon based therapy.

Materials And Methods

Retrieval of Whole Genomes and Conserved Region:

Public database Global Initiative on Sharing All Influenza Data (GISAID) was explored to retrieve the whole-genomes of novel SARS CoV–2[18]. Sixty-seven (67) whole-genomes of SARS CoV–2 were selected for this present study (January, 2020). In quest for developing universal therapeutics, multiple sequence alignment and construction of phylogenetic cladogram were performed using neighbor joining method in CLC Drug Discovery Workbench (Version 3.0). For further study, consensus sequences were collected to develop a vaccine, small molecules and SiRNAs.

Chimeric Vaccine against SARS-CoV–2

Antigenicity analysis of the candidate proteins

The antigenicity of the selected Membrane Glycoprotein (NCBI accession:YP_009724393.1) and Spike Glycoprotein (NCBI accession: QHR63250.2) were determined by Vaxijen 2.0[18] and AntigenPro[19]. These servers predicted antigenicity through alignment independent methods.

Identification of B cell and T cell Epitopes

The outside amino acid residues of the vaccine candidates were identified with THMM[19]. The identified sequences were uploaded in BepiPred–2.0 server to select the most potential B cell epitopes[21]. During this selection, surface exposed amino acids were prioritized mostly. The Cytotoxic T cell (CTL), Helper T cell (HTL) epitopes were disclosed by NetCTL 1.2 and NetMHC 4.0 respectively[22,23]. NetCTL 1.2 was implemented to discover the CTL epitopes for 12 classes of Major Histocompatibility Complex 1 (MHC I) supertypes. The best CTL epitopes were selected based on combined score. NetMHCII 2.2 was used to detect 15-mer HTL epitopes for human HLA-DP, HLA-DQ, and HLA-DR alleles. Most potential epitopes were chosen by evaluating affinity, percentage ranking and binding level.

Construction of Chimeric Vaccine

To construct a chimeric or multi-epitope vaccine, all the epitopes were joined with EAAK, GPGPG and AAY linkers[24]. Human Beta Defensin–2 (HBD–2) (PDB ID: 1FD3) and a recombinant viral protein were added in the N terminal and the C terminal of the vaccine respectively[25,26]. HBD–2 was conjugated because the protein can activate Toll Like Receptor 4 (TLR 4) and have chemotactic activity[25,27]. The recombinant viral protein was added to stimulate the antiviral responses[26].

Evaluation of Immune Response and Interferon Gamma (IFNγ)

The fasta sequence of the vaccine was uploaded to an agent-based immune system simulator C-ImmSim (http://150.146.2.1/C-IMMSIM/index.php) for measuring the immune responses[28]. The parameters were kept default during this simulation. C-ImmSim showed an adequate secretion of IFNγ which was further evaluated by IFNepitope[29].

Allergenicity and Toxicity Exploration

Recombinant vaccine can initiate allergic response or lead to various types of toxicity. Therefore, evaluation of toxicity and allergenicity is an essential step for vaccine design. Allergenicity of the vaccine was calculated by AlgPred and AllerTop v.2[30,31]. Toxicity was measured by ToxinPred[32].

Analysis of Physicochemical Properties and Tertiary Structure

The physicochemical analysis of the protein was executed via ProtParam[32]. The tertiary structure was generated through Contact-guided Iterative Threading ASSEmbly Refinement (C-I-TASSER) (17). The structure was refined by GalaxyRefine[35] and validated with Procheck[36] and ProSAWeb[37].

Molecular Docking Analysis

The crustal structure of TLR 4 was collected from Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) (www.rcsb.org). The PDB ID of the structure was 3FXI. For molecular docking analysis, only Chain A of TLR 4 was preserved and prepared by deleting the heterogeneous atoms. This preparation was executed by BIOVIA Discovery Studio (https://www.3dsbiovia.com/). To determine the active site pocket of Chain A, the PDB file was uploaded in Computer Atlas of Surface Topology of Proteins (CASTp)[38]. Finally, the vaccine and Chain A of TLR–4 was uploaded in ClusPro 2.0 for protein-protein docking[39].

Codon Optimization and Visualization of Cloning

The vaccine sequence was reversely translated and the codons were optimized for *Escherichia coli* (strain K12) through JCat[40]. The optimized DNA sequence was kept free from rho-independent transcription terminators and prokaryotic ribosome binding sites. This sequence was flanked by Nde I and Xho I restriction sites. Stop codons were added at the end of 3'OH or C terminal end. After these preparations, the DNA sequence was inserted in pET28a (+) Plasmid Vector via SnapGene tools (www.snapgene.com).

Small Molecule Therapeutics against COVID–19

Homology Modelling and Binding site analysis

The conserved RNA directed RNA polymerase (RdRp) sequence was retrieve to develop small molecules. C-I-TASSER was employed to build the 3D model of RdRP[34]. Thereafter, CASTp was applied to identify the drug binding sites that allowed to recognize the critical amino acids for drug-protein interactions[38].

Screening the druggable compounds and molecular docking simulation

DrugBank Database (www.drugbank.ca)was utilized to search for the potent drugs against RdRP. Molecular docking of receptorand the selected compounds were performed in AutodockVina to observe the binding affinity into the binding site of RdRp[42]. Here, AutoDock tools 1.5.6 was used to prepare the input pdbqt file for the receptor. A grid box parameter were set in size with X = 70, Y = 70, Z = 36 points (center grid box:x = 23.448, y = 53.587, z = –2.012, spacing = 0.5˚A). The molecular visualization of protein-ligands were analyzed by BIOVIA Discovery Studio and PyMol[43].

Pharmacoinformatics illustration

Osiris property explorer[44], Molinspiration[45], ACToR (Aggregated Computational Toxicology Resource)[46], admetSAR (absorption, distribution, metabolism, excretion, and toxicity Structure-Activity Relationship database)[47] and ACD/I-lab[48] were exploited for the calculation of Absorption, distribution, metabolism, excretion (ADME) properties and toxicity profile. ADMET properties are necessary to establish an effective drug.

Nucleic Acid Based Therapeutics Development

Designing of potential siRNA molecules

In order to design the effective siRNA molecule, I-Score Designer was employed[49]. I-Score Designer computes nine different siRNA designing scores such as Ui-Tei[50], Amarzguioui[51], Hsieh[52], Takasaki[53], s-Biopredsi[54], i-Score[55], Reynolds[56], Katoh[57], and Design of SIRna (DSIR) [58] along with other essential parameters. The server also ranks the best siRNA molecules. From there, the best siRNA sequence was taken for further analysis. The secondary structure of the siRNA was predicted via RNA structure webserver[59]. RNAfold web server was applied to calculate the free energy of the thermodynamic ensemble for the secondary structures[60]. Transcription and Translation Tool (http://biomodel.uah.es/en/lab/cybertory/analysis/trans.htm) was implemented to transcribe the viral DNA sequences. Finally, the designated Antisense siRNAs were hybridized against viral RNA sequences using the HNADOCK server[61]. HNADOCK executed RNA-RNA docking and performed molecular dynamics simulations to refine the best 10 predicted siRNA-mRNA complexes. The model with best docking score for Membrane Glycoprotein mRNA and Spike Glycoprotein mRNA were visualized with PyMol.

Inteferon stimulating genes (ISGs) based Interferon Therapy

Comparative Genomics

The genome of SARS Cov–2 was compared with SARS CoV and MERS CoV. We Blasted SARS CoV–2 against SARS CoV and MERS CoV, using Megablast and Discontiguous Megablast algorithm. The graphical representation of side by side genome comparison is demonstrated in Artemis Comparison Tool[62].

Exploring SARS-CoV expression profile

The Gene Expression profile (GSE5972) of SARS-CoV was collected from National center for Biotechnology Information (NCBI). Afterwards, the normalization study was performed in between 10 Healthy samples and 54 SARS patient samples, whereas recovered cases were excluded from this study. We used limma R package to identify the differentially expressed genes[63]. IDEP tools was utilized to create Hierarchical Clustering Heatmap and Boxplot to the visualization and distribution of the data for both up and downregulated genes[64].

Gene Enrichment Analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) and Enrichr4 were employed to identify the genes which were involved in pathways and Gene Ontology (GO) processes such as biological process (BP), molecular function (MF) and cellular component (CC)[65–67]. The genes with GO accession ID &KEGG pathways were enlisted for further study. ShinyGO was utilized to construct a clustered tree of the top 30 GO terms for both up and downregulated genes of BP, MF and CC[68].

Identification of Viral Genes

We securitized the down and upregulated genes in BP, MF and CC from gene ontology dataset to find out the viral connected genes especially enriched viral production regulation and cytokine regulation.

Exploration of Interferon Stimulating Genes (ISGs)

INTERFEROME[69] database was analyzed to recognize the Interferon Stimulating Genes (ISGs) and potential Interferons. Further, we tried to explore the pathway of ISGs which modulate the immune system and interferon regulation by Reactome[70].

Results

Sixty seven (67) SARS-CoV-2 whole genome strains were collected from different countries (Supplementary Data 1). Multiple Sequence Alignment (MSA) and Phylogenetic Tree depicted the similarity and distant relationship of SARS-CoV-2 among the different geographic area and populations (Supplementary Data 2 and Supplementary Fig. 1). From these analysis, 3 consensus sequences from the whole genomes were collected (Supplementary Data 3). The conserved sequences were translated and their functions were analyzed manually to develop vaccine, siRNAs and drugs.

Chimeric Vaccine against SARS-CoV-2

The selected membrane glycoprotein (MG) (Vaxijen score 0.55 and AntigenPro score 0.232547) and spike glycoprotein (SG) (Vaxijen score 0.4695 and AntigenPro score 0.717053) were predicted as “antigenic” by the assigned programs. The MG and SG have a total of 24 and 1214 outside amino acids respectively. Among those residues, one potential B cell epitope NGTITVEELKKLLEQ was identified in MG (Table 1). Four possible B-cell epitopes (HAPATVCGPKKSTN, NNSYECDIP, FYEPQIITTD, VEGFNCYFPLQ) from the SG and only one from the MG (NGTITVEELKKLLEQ) were found (Table 1). The selected four Cytotoxic T-cell (CTL) epitopes from MG and SG cover HLA-A12, A24, A26, B58, B8 alleles. Moreover, Helper T-cell (HTL) epitopes can interact with HLA DRB1-0101, DPA10103-DPB10301, DQA10101-DQB10501 and DRB1-1501 (Supplementary Table 1). The chimeric vaccine exposed the strong binding affinity with Toll Like Receptor-4 (TLR4) (Fig. 2A). A single injection of the vaccine has showed an adequate release of IFNγ in C-ImmSim simulator (Fig. 2B). The IFNepitope also showed 127 IFNγ inducing epitopes by different methods. Additionally, Interleukin (IL)-2 was also released significantly for lymphocyte differentiation. This differentiations are mentioned as danger signal D (Sympshon Index) (Fig. 2B). After release of inflammatory cytokines, the immune system released anti-inflammatory IL-10 at 2.5th day and consequently generated Immunoglobulin M (IgM) and later IgG (Fig. 2C). Generation of memory B cells was also

observed (Fig. 2D). The vaccine specifically stimulated the differentiation of T helper Cell- 1 (Th 1) with certain level of T regulatory cell (Fig. 2E). Memory Helper T Cell also released in between 2.5 days (Figs. 2F and 2G). The vaccine does not contain any major allergenic properties according to AlgPred and AllerTop. ToxinPred identified 343 non-toxic peptides in different combinations and only 11 peptides with toxic properties. The vaccine is stable according to the instability index of ProtParam (Supplementary Table 2). The estimated half-life of the protein is 30 hours in mammalian reticulocytes (*in vitro*), >20 hours in yeast (*in vivo*) and >10 hours in *Escherichia coli* (*in vivo*) with -0.010 Grand Average of Hydropathicity (GRAVY) and 81.38 Aliphatic index (Supplementary Table 2). The refined three dimensional (3D) structure of the vaccine has 1.905 MolProbity score (provided by GalaxyWEB), -2.62 Z-Score and 87.9% residues in the most favored regions of Ramachandran Plot (Supplementary Figs. 2 and 3). When this protein was docked against TLR 4 (Chain A) via ClusPro 2.0, it interacted toward the receptor via active site pocket and HBD-2 with lowest energy score -1401.1.

Table 1: The potential B cell, Cytotoxic T cell and Helper T cell epitopes.

Type of Epitope	Serial No.	Epitope Sequence	Protein Name
B Cell epitopes	1	NGTITVEELKKLLEQ	Membrane Glycoprotein
	2	HAPATVCGPKKSTN	Spike Glycoprotein
	3	NNSYECDIP	Spike Glycoprotein
	4	FYEPQIITTD	Spike Glycoprotein
	5	VEGFNCYFPLQ	Spike Glycoprotein
CTL Epitopes	6	SSDNIALLV	Membrane Glycoprotein
	7	YFIASFRLF	Membrane Glycoprotein
	8	YIIKLIFLW	Membrane Glycoprotein
	9	LAAVYRINW	Membrane Glycoprotein
	10	LTDEMIAQY	Spike Glycoprotein
	11	FVFKNIDGY	Spike Glycoprotein
	12	YLQPRTFLL	Spike Glycoprotein
	13	RSFIEDLLF	Spike Glycoprotein
HTL Epitopes	14	TLSYYKLGASQRVAG	Membrane Glycoprotein
	15	ASFRLFARTRSMWSF	Membrane Glycoprotein
	16	SNLLLQYGSFCTQLN	Spike Glycoprotein
	17	LIRAAEIRASANLAA	Spike Glycoprotein
	18	WFVTQRNFYEPQIIT	Spike Glycoprotein

Codon Optimization and Cloning for chimeric vaccine

The DNA sequence of the vaccine has 1089 nucleotides that demonstrated 0.96 Codon Adaptation Index (CAI) with 54.26% GC content. Here, the GC content of the host organism *E. coli* strain K12 is 50.73 (Fig. 3). This sequence was taken under modifications for cloning

Small Molecule Drugs against COVID-19

Six experimental compounds were collected from the DrugBank. The chemical properties of the compounds showed the possibility to be suitable candidates against the RdRp (Supplementary Table 3). Further, the target protein and six compounds were selected for Molecular docking analysis to ensure the binding affinity. Strong binding affinity was observed for benzyl (2-oxopropyl)carbamate (-4.2 Kcal/mol) (Fig. 4A), 2-[(2,4-dichloro-5-methylphenyl)sulfonyl]-1,3-dinitro-5-(trifluoromethyl)benzene (-7.6 Kcal/mol) (Fig. 4B),; s-[5-(trifluoromethyl)-4h-1,2,4-triazol-3-yl]5-(phenylethynyl)furan-2-carbothioate (-5.9 Kcal/mol) (Fig. 4C), 5-amino-2-methyl-N-[(1R)-1-naphthalen-1-ylethyl]benzamide (-6.3 Kcal/mol) (Fig. 4D), Nalpha-[(benzyloxy)carbonyl]-n-[(1r)-4-hydroxy-1-methyl-2-oxobutyl]-l-phenylalaninamide (-7.4 Kcal/mol) (Fig. 4E), 4-(Dimethylamino)benzoic acid (-4.0 Kcal/mol) (Fig. 4F). Fifteen common amino acid residues including Val473, Glu474, Asp477, Lys478, Leu636, Val637, Arg735, Asp736, Val737, Asp304, Arg305, Arg640, His642, Arg651 and Leu648 of RdRp interacted with the druggable molecules. Later, ADMET analysis of the experimental compounds were enlisted (Supplementary Table 4).

Silencing the glycoproteins of SARS-CoV-2

The Sense and Antisense RNA molecules of both MG and SG were 19mer and 21mer respectively (Supplementary Table 5 and 6). The siRNA generated hairpin like secondary structure at 310.15 K temperature. The free energy of the thermodynamic ensembles were -18.30 kcal/mol and -17.90 kcal/mol respectively for MG and SG siRNA secondary structures. Results of HNADOCK server demonstrate that the antisense siRNAs can interact with the three dimensional (3D) viral mRNA molecules with around -260 docking score (Fig. 5).

Interferon Based Therapeutic for SARS-CoV-2

The whole genome of SARS-CoV and MERS-CoV were compared with SARS-CoV-2. SARS-CoV and SARS-CoV-2 share 88% query sequence coverage with 82.30% sequence identity. Besides, MERS-CoV share only 29% query coverage with 67.06% sequence identity (Supplementary Fig. 4). Genome alignment showed higher homology between SARS-CoV and SARS CoV-2 with a very few synteny break region. Whereas, MERS-CoV shares low scoring homology with SARS-CoV-2 with a larger synteny breakage area. The entry of SARS-CoV-2 and SARS-CoV are similar so differential gene expression investigation was run on SARS-CoV (Fig. 6).

64 different samples were normalized and median was centralized (Fig. 7A). 19,200 differentially expressed genes were found (Fig. 7B and Supplementary Fig. 5). Genes with official human gene symbol (HGNC) and have a significant p-value (<0.05) were selected (Supplementary Data 4 and 5). 1265 genes were considered for further analysis, among them 616 genes were upregulated (Supplementary Data 6) and 649 genes were downregulated (Supplementary Data 7). In the hierarchical clustering heatmap genes based on LogFC, t-stat and adjusted p-value were clustered together (Figs. 7C, D, E and F). The genes that we have considered as upregulated seem to have major frequency around LogFC≈ 2.1 and t value≈ 2.8, which dictates significant upregulation (Supplementary Data 8). On the other side, the downregulated genes seem to have major frequency around LogFC≈ -1.7 and t value≈ -2.4 (Supplementary Data 9). 582 upregulated and 579 downregulated genes were mapped through Gene Ontology and KEGG Pathway (Supplementary Data 8 and 9). For overexpression, 231 BP (Supplementary Data 10), 61 MF (Supplementary Data 11) and 44 CC (Supplementary Data 12) genes with their GO accession and 77 KEGG human pathways (Supplementary Data 13) were enlisted. For underexpression, 97 BP (Supplementary Data

14), 60 MF (Supplementary Data 15) and 54 CC (Supplementary Data 16) genes with GO accession and 101 KEGG pathways (Supplementary Data 17) were enlisted as well.

The biological process of Gene Ontology (GO) reported Upregulated 91 genes and Downregulated 76 genes to be connected with viral regulation process (Supplementary Data 18, 19, 20 and 21). The upregulated (KPNA4, IRF7, OAS3, CAMK2G, IFITM1, IFIT1, EGR1, MX1 and IFNAR2) and the downregulated (TRIM8, VCAM1, IFNGR2 and JAK1) genes were then identified to be effective to stimulate the Interferons via the pathway regulation (Fig. 8, Supplementary Fig. 6, and Supplementary Data 21, 22, 23, 24 and 25).

Discussion

COVID-19 has become a pandemic disease and causing severe trouble in social, political and economic stability[71,72]. The causative agent of this pandemic can go under frequent mutations, adaptations and transmission that possess continuous risk of outbreaks[73,74]. To handle the current situation and future risks, sustainable development of therapeutics against this deadly virus is a crying need. To address this global issue, we have merged several computational approaches and proposed therapeutics that could be effective against all the strains of SARS CoV-2 (Fig 1). To achieve this goal, all available whole genomes of SARS-CoV-2 from different countries were exploited (Supplementary Data 1). Among all the strains, several consensus region was obtained with minimal evolutionary distance (Supplementary Data 2, 3 and Supplementary Fig 1). The predicted chimeric vaccine, drug compounds and siRNA based therapeutics were proposed on the basis of those conserved sequences (Supplementary Data 3).

Two proteins namely Membrane and Spike Glycoprotein were identified through blastx of the conserved nucleotide sequence. The vaccine candidates contain several well-conserved B and T cell epitopes (Table 1, Supplementary Table 1). The epitopes were exploited to design a potential chimeric vaccine which is expected to elicit TLR4 mediated Th_1 antiviral response. Simulation study also showing the ability of this vaccine is producing memory B and T cell along with the protective immune-globulins (Figs. 2 and 3). The vaccine is expected to induce inflammatory and anti-inflammatory cytokines in sufficient amounts with non-allergen activities. Finally, since the vaccine contains very few toxic peptides; therefore, *in vitro* and *in vivo* validation are required prior to the actual application. Again, the constructed expression vector should successfully produce adequate vaccine molecules in *E. coli* strain K12 since the half-life of the protein is >10 hours inside the bacterial host with an acceptable stability index (Fig. 3).

A universal drug is necessary with an effective vaccine to reduce the post infection consequences of COVID-19. For coronaviruses, RNA-directed RNA polymerase (RdRp) is an attractive therapeutic target which catalyzes the replication of RNA from RNA template. Thus, we targeted the conserved (region 256nt-13458nt) RNA directed RNA polymerase (RdRP) protein from SARS-CoV-2 (Supplementary Data 3). The virtual screening directed by the RdRP protein identified six experimental compounds from the DrugBank server (Supplementary Table 3). Later, the molecular docking analysis revealed a higher binding affinity between the compounds and the RdRp receptor. It is important that all the interacting residues between drug compounds and RdRp receptor are located in the predicted drug binding site (Fig. 4). Among the ADMET properties, all the experimental compounds exhibited the sufficient human intestinal absorption rate, a higher metabolic rate, no blood-brain barrier permeability, and balanced amounts of drug supply to the body (Supplementay Table 4). Moreover, the toxicity levels of these compounds appeared to be safe for the human body (Supplementay Table 4). These characteristics are

significant for drug development because poor ADMET properties could direct a drug to be failing in phase III clinical trials[75].

RNA interference (RNAi) pathway is very efficient and specific in silencing the viral genes just after post-transcription. Here, siRNAs (small interfering RNAs) can play a therapeutic role[76,77]. The mRNA sequences have potential sense siRNA strands CAGUAUAAUUAAUAACUAA and GGUUUUUGUAUAUAAUUAA from SG and MG respectively (Supplementary Table 5 and 6). Therefore, 19mer and 21mer nucleotide of sense and antisense strand were designed on the basis of siRNA universal rules. The standard thermodynamic value and strong interaction between antisense and viral mRNA strongly supports its prospective application (Figs. 5A and 5B). These siRNAs might be conjugated with aptamer or delivered through various delivery systems to inhibit the intracellular growth of SARS-CoV-2.

Interferons (IFNs) instigate an array of antiviral effectors. Every virus contains a distinctive property, however, partly overlapping antiviral Interferon Stimulating Genes or "ISG profile" could be potential target. IFN effectors target various phases of the virus life cycle to interfere with viral infection and proliferation. To identify the ISG and IFN of COVID-19, the gene expression profile of COVID-19 patients was needed. Due to a lack of the expression profile of COVID-19 till January 2020, three genomes of Betacoronavirus were taken including SARS-CoV, MERS-CoV and SARS-CoV-2, which have common virulence patterns, sign and symptoms. A comparison of their genome was performed. The whole genome comparison between SARS-CoV and SARS-CoV-2 showed 88% similarity with each other (Fig. 6 and Supplementary Fig 4). Therefore, the SARS CoV expression profile was explored for the identification of ISG and IFNs for the IFNs based therapeutics development. On the basis of this connection, we analyzed the genome-wide expression profile of SARS-CoV to identify the ISG and IFNs which might be suitable candidates for the IFNs based therapeutic development (Supplementary Data 4, 5, 6, 7, 8 and 9). Among the up and down regulated genes, a total of 167 immunoregulatory genes (Upregulated 91 genes and Downregulated 76 genes) were found from the biological process of Gene Ontology (GO) (Fig. 7, Supplementary Fig. 5, Supplementary Data 9, 10, 11, 12, 13, 14, 15, 16 and 17). We then identified the Interferon stimulating genes KPNA4, IRF7, OAS3, CAMK2G, IFITM1, IFIT1, EGR1, MX1 and IFNAR2 from the upregulated immunoregulatory genes and TRIM8, VCAM1, IFNGR2, JAK1 from the downregulated immunoregulatory genes (Fig.8). These immunoregulatory genes can prompt the interferon IFN alpha, IFN beta and IFN gamma which then directly gets involved to induce the immunological response by regulating the pathways (Fig. 8, Supplementary Fig. 6, Supplementary Data 17, 18, 19, 20, 21, 22, 23, 24 and 25). This analysis gives a deep insight for the IFN based therapy during COVID-19.

The pandemic coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is of great global public health concern due to its spreading rate in every part of this world with increasing death rate that ultimately has led to global socioeconomic disruption. The global urgency for the antiviral therapy against the SARS-CoV-2 has been the superior priority of this world to survive the existence of human as no medication is available so far. The present study proposed all possible novel therapeutics such as chimeric vaccine, potent drug compounds, effective siRNA and interferons to the purpose of shortening the time and reduce the cost for effective medication discovery against the global pandemic COVID-19. With the aid of computational biology and Bioinformatics analyses, these therapeutics confirmed its effectiveness against the COVID-19 which further offer for the wet lab validation. Lastly, we systematically proposed the possible therapeutics in the hope of providing a reference to minimize the longer time required antiviral therapy in wet lab and help for the prevention as well as control of the COVID-19 pandemic.

Conclusion

The present study has proposed four types of therapeutics including chimeric vaccine, potent drugs, siRNAs and ISG targets against COVID–19. We have analyzed all the genomic, transcriptomic and proteomic data to anticipate the possible therapeutics against the recent pandemic COVID–19. Therefore, these findings might be of interest to the researchers and pharmaceuticals company to discover the novel therapeutics against the SARS-CoV–2. We hope this *in silico* pipeline will also help many researchers to combat against other viral outbreaks in future.

Declarations

Acknowledgements

The Authors are grateful to Ministry of Science and Technology for the extensive support during this work.

Author Contributions

MS: Conceived, designed, and guided the study, analyzed the data, drafting the manuscript and performed critical revision. CAK and KCD: Guided the study, acquisition and analyzed the data, helped in drafting the manuscript. MM and MGM: Helped in Bioinformatics analysis and helped in drafting the manuscript. MUH, AB, MTHE, ZMC: Helped to design the study, performed bioinformatics analysis, drafted, and developed the manuscript and performed critical revision. All authors have approved the manuscript.

Competing Interest

The authors declare that they have competing interest on proposed therapeutics.

Funding

There is no funding source for this work

References

1. WHO. Naming the coronavirus disease (COVID-19) and the virus that causes it 2029 (2019).
2. Lai, C.C., Shih, T.P., Ko, W.C., Tang, H.J. and Hsueh, P.R., Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. International journal of antimicrobial agents, p.10592 (2020).
3. Wu, Y.C., Chen, C.S. and Chan, Y.J., The outbreak of COVID-19: An overview. Journal of the Chinese Medical Association, 83(3), pp.217-220 (2020).
4. Fung et al. Human Coronavirus: Host-Pathogen Interaction. *Annual review of microbiology*, *73*, pp.529-557 (2019).
5. Cavanagh D. Coronaviridae: a review of coronaviruses and toroviruses. InCoronaviruses with Special Emphasis on First Insights Concerning SARS (pp. 1-54). Birkhäuser Basel(2005).
6. Cui, J., Li, F. & Shi, Z. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 17, 181–192 (2019).

7. Lau, S.K., Wong, A.C., Zhang, L., Luk, H.K., Kwok, J.S., Ahmed, S.S., Cai, J.P., Zhao, P.S., Teng, J.L., Tsui, S.K. and Yuen, K.Y., Novel Bat Alphacoronaviruses in Southern China Support Chinese Horseshoe Bats as an Important Reservoir for Potential Novel Coronaviruses. *Viruses*, *11*(5), p.423 (2019).
8. Fan, Y., Zhao, K., Shi, Z.L. and Zhou, P.,. Bat Coronaviruses in China. *Viruses*, *11*(3), p.210 (2019).
9. Wu, Z. and McGoogan, J.M., 2020. Characteristics of and important lessons from the coronavirus disease (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *Jama* (2019).
10. Tang, A., Tong, Z.D., Wang, H.L., Dai, Y.X., Li, K.F., Liu, J.N., Wu, W.J., Yuan, C., Yu, M.L., Li, P. and Yan, J.B., Detection of Novel Coronavirus by RT-PCR in Stool Specimen from Asymptomatic Child, China. *Emerging infectious diseases*, *26*(6) (2020).
11. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A. and Müller, M.A., SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell (2020).
12. Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y. and Yu, T., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet, 395(10223), pp.507-513 (2020).
13. Gao, J., Tian, Z. and Yang, X., Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. BioScience Trends (2020).
14. Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S.C. and Di Napoli, R., Features, Evaluation and Treatment Coronavirus (COVID-19). In StatPearls [Internet]. StatPearls Publishing (2020).
15. Hossain, M.U., Omar, T.M., Alam, I., Das, K.C., Mohiuddin, A.K.M., Keya, C.A. and Salimullah, M., Pathway based therapeutic targets identification and development of an interactive database CampyNIBase of Campylobacter jejuni RM1221 through non-redundant protein dataset. PloS one, 13(6) (2018).
16. Abadio, A.K.R., Kioshima, E.S., Teixeira, M.M., Martins, N.F., Maigret, B. and Felipe, M.S.S., Comparative genomics allowed the identification of drug targets against human fungal pathogens. *BMC genomics*, *12*(1), p.75 (2011).
17. Bhattacharjee, A., Hossain, M.U., Chowdhury, Z.M., Rahman, S.A., Bhuyan, Z.A., Salimullah, M. and Keya, C.A., Insight of Druggable Cannabinoids against Estrogen Receptor β in Breast Cancer. *Journal of Biomolecular Structure and Dynamics*, (just-accepted), pp.1-12 (2020).
18. Doytchinova, I.A. and Flower, D.R., Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine*, *25*(5), pp.856-866 (2007).
19. Magnan, C.N., Zeller, M., Kayala, M.A., Vigil, A., Randall, A., Felgner, P.L. and Baldi, P., High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics*, *26*(23), pp.2936-2943 (2010).
20. Krogh, A., Larsson, B., Von Heijne, G. and Sonnhammer, E.L., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, *305*(3), pp.567-580 (2001).
21. Jespersen, M.C., Peters, B., Nielsen, M. and Marcatili, P., BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic acids research*, *45*(W1), pp.W24-W29 (2017).
22. Larsen, M.V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O. and Nielsen, M., Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics*, *8*(1), p.424 (2007).

23. Andreatta, M. and Nielsen, M., Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*, *32*(4), pp.511-517 (2016).
24. Shey, R.A., Ghogomu, S.M., Esoh, K.K., Nebangwa, N.D., Shintouo, C.M., Nongley, N.F., Asa, B.F., Ngale, F.N., Vanhamme, L. and Souopgui, J., In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Scientific reports*, *9*(1), pp.1-18 (2019).
25. Yu, L., Wang, L. and Chen, S., Endogenous toll-like receptor ligands and their biological significance. *Journal of cellular and molecular medicine*, *14*(11), pp.2592-2603 (2010).
26. Li, J., Ulitzky, L., Silberstein, E., Taylor, D.R. and Viscidi, R., Immunogenicity and protection efficacy of monomeric and trimeric recombinant SARS coronavirus spike protein subunit vaccine candidates. *Viral immunology*, *26*(2), pp.126-132 (2013).
27. Niyonsaba, F., Ogawa, H. and Nagaoka, I., Human β-defensin-2 functions as a chemotactic agent for tumour necrosis factor-α-treated human neutrophils. *Immunology*, *111*(3), pp.273-281 (2004).
28. Rapin, N., Lund, O., Bernaschi, M. and Castiglione, F., Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One*, *5*(4) (2010).
29. Dhanda, S.K., Vir, P. and Raghava, G.P., Designing of interferon-gamma inducing MHC class-II binders. *Biology direct*, *8*(1), p.30 (2013).
30. Saha, S. and Raghava, G.P.S., AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic acids research*, *34*(suppl_2), pp.W202-W209 (2006).
31. Dimitrov, I., Bangov, I., Flower, D.R. and Doytchinova, I., AllerTOP v. 2—a server for in silico prediction of allergens. *Journal of molecular modeling*, *20*(6), p.2278 (2014).
32. Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Raghava, G.P. and Open Source Drug Discovery Consortium, In silico approach for predicting toxicity of peptides and proteins. *PloS one*, *8*(9) (2013).
33. Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D. and Bairoch, A., Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571-607). Humana press (2005).
34. Zheng, W., Li, Y., Zhang, C., Pearce, R., Mortuza, S.M. and Zhang, Y., Deep-learning contact-map guided protein structure prediction in CASP13. *Proteins: Structure, Function, and Bioinformatics*, *87*(12), pp.1149-1164 (2019).
35. Heo, L., Lee, H. and Seok, C., GalaxyRefineComplex: Refinement of protein-protein complex model structures driven by interface repacking. *Scientific reports*, *6*, p.32153 (2016).
36. Laskowski, R.A., MacArthur, M.W. and Thornton, J.M., PROCHECK: validation of protein-structure coordinates (2006).
37. Wiederstein, M. and Sippl, M.J., ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, *35*(suppl_2), pp.W407-W410 (2007).
38. Tian, W., Chen, C., Lei, X., Zhao, J. and Liang, J., CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic acids research*, *46*(W1), pp.W363-W367 (2018).
39. Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D. and Vajda, S., The ClusPro web server for protein–protein docking. *Nature protocols*, *12*(2), p.255 (2017).
40. Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D.C. and Jahn, D., JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic acids research*, *33*(suppl_2), pp.W526-

W531 (2005).

41. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B. and Jain, S., MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*, *27*(1), pp.293-315 (2018).
42. Agrawal, S., Gupta, D. and Panda, S.K., The 3' end of hepatitis E virus (HEV) genome binds specifically to the viral RNA-dependent RNA polymerase (RdRp). Virology, 282(1), pp.87-101 (2001).
43. Seeliger, D. and de Groot, B.L., Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of computer-aided molecular design*, *24*(5), pp.417-422 (2010).
44. Sander, T., OSIRIS property explorer. *Organic Chemistry Portal* (2001).
45. Kumar Mishra, N. and PS Raghava, G., Prediction of specificity and cross-reactivity of kinase inhibitors. *Letters in Drug Design & Discovery*, *8*(3), pp.223-228 (2011).
46. Judson, R., Richard, A., Dix, D., Houck, K., Elloumi, F., Martin, M., Cathey, T., Transue, T.R., Spencer, R. and Wolf, M., ACToR—aggregated computational toxicology resource. Toxicology and applied pharmacology, 233(1), pp.7-13 (2008).
47. Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z., Liu, G., Lee, P.W. and Tang, Y., admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties (2012).
48. Masunov, A., ACD/I-Lab 4.5: an internet service review. Journal of chemical information and computer sciences, 41(4), pp.1093-1095 (2001).
49. Ichihara, M., Murakumo, Y., Masuda, A., Matsuura, T., Asai, N., Jijiwa, M., Ishida, M., Shinmi, J., Yatsuya, H., Qiao, S. and Takahashi, M., Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. *Nucleic acids research*, *35*(18), p.e123 (2007).
50. Ui-Tei, K., Naito, Y., Zenno, S., Nishi, K., Yamato, K., Takahashi, F., Juni, A. and Saigo, K., Functional dissection of siRNA sequence by systematic DNA substitution: modified siRNA with a DNA seed arm is a powerful tool for mammalian gene silencing with significantly reduced off-target effect. *Nucleic acids research*, *36*(7), pp.2136-2151 (2008).
51. Amarzguioui, M. and Prydz, H., An algorithm for selection of functional siRNA sequences. *Biochemical and biophysical research communications*, *316*(4), pp.1050-1058 (2004).
52. Hsieh, A.C., Bo, R., Manola, J., Vazquez, F., Bare, O., Khvorova, A., Scaringe, S. and Sellers, W.R., A library of sirna duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic acids research*, *32*(3), pp.893-901 (2004).
53. Takasaki, S., Kotani, S. and Konagaya, A., An effective method for selecting siRNA target sequences in mammalian cells. *Cell cycle*, *3*(6), pp.788-793 (2004).
54. Huesken, D., Lange, J., Mickanin, C., Weiler, J., Asselbergs, F., Warner, J., Meloon, B., Engel, S., Rosenberg, A., Cohen, D. and Labow, M., Design of a genome-wide siRNA library using an artificial neural network. *Nature biotechnology*, *23*(8), pp.995-1001 (2005).
55. Ichihara, M., Murakumo, Y., Masuda, A., Matsuura, T., Asai, N., Jijiwa, M., Ishida, M., Shinmi, J., Yatsuya, H., Qiao, S. and Takahashi, M., Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. *Nucleic acids research*, *35*(18), p.e123 (2007).
56. Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S. and Khvorova, A., Rational siRNA design for RNA interference. *Nature biotechnology*, *22*(3), pp.326-330 (2004).

57. Katoh, T. and Suzuki, T., Specific residues at every third position of siRNA shape its efficient RNAi activity. Nucleic acids research, 35(4), p.e27 (2007).
58. Vert, J.P., Foveau, N., Lajaunie, C. and Vandenbrouck, Y., An accurate and interpretable model for siRNA efficacy prediction. BMC bioinformatics, 7(1), p.520 (2006).
59. Mathews, D.H. and Turner, D.H., Prediction of RNA secondary structure by free energy minimization. Current opinion in structural biology, 16(3), pp.270-278 (2006).
60. Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M. and Turner, D.H., Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proceedings of the National Academy of Sciences, 101(19), pp.7287-7292 (2004).
61. He, J., Wang, J., Tao, H., Xiao, Y. and Huang, S.Y., HNADOCK: a nucleic acid docking server for modeling RNA/DNA–RNA/DNA 3D complex structures. Nucleic acids research, 47(W1), pp.W35-W42 (2019).
62. Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., Parkhill, J., ACT: the Artemis Comparison Tool. Bioinformatics, 21(16):3422-3 (2005).
63. Ritchie, M.E., Phipson, B., Wu, D.I., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K., limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, *43*(7), pp.e47-e47 (2015).
64. Ge, S.X., Son, E.W. and Yao, R., iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC bioinformatics*, *19*(1), p.534 (2018).
65. Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C. and Lempicki, R.A., The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome biology*, *8*(9), p.R183 (2007).
66. Kanehisa, M. and Goto, S., KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), pp.27-30 (2000).
67. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A. and McDermott, M.G., Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, *44*(W1), pp.W90-W97 (2016).
68. Ge, S. and Jung, D., ShinyGO: a graphical enrichment tool for animals and plants. *Biorxiv*, p.315150 (2018).
69. Samarajiwa, S.A., Forster, S., Auchettl, K. and Hertzog, P.J., INTERFEROME: the database of interferon regulated genes. Nucleic acids research, 37(suppl_1), pp.D852-D857 (2009).
70. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R. and Loney, F., The reactome pathway knowledgebase. Nucleic acids research, 48(D1), pp.D498-D503 (2020).
71. Ren, S. Y., Gao, R. D., & Chen, Y. L., Fear can be more harmful than the severe acute respiratory syndrome coronavirus 2 in controlling the corona virus disease 2019 epidemic. World journal of clinical cases, 8(4), 652–657 (2020).
72. Berera, D., & Zambon, M., Antivirals in the 2009 pandemic–lessons and implications for future strategies. Influenza and other respiratory viruses, 7 Suppl 3(Suppl 3), 72–79. (2013).
73. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C. and Garry, R.F., The proximal origin of SARS-CoV-2. *Nature Medicine*, pp.1-3 (2020).

74. Gurwitz, D., Angiotensin receptor blockers as tentative SARS-CoV-2 therapeutics. Drug Development Research (2020).
75. Ju, S., Tardiff, D.F., Han, H., Divya, K., Zhong, Q., Maquat, L.E., Bosco, D.A., Hayward, L.J., Brown, R.H., Lindquist, S., Ringe, D., Petsko, G.A., A yeast model of FUS/TLS-dependent cytotoxicity. PLoS Biol, Apr;9(4):e1001052 (2011)
76. Fire, A., Xu, S., Montgomery, M.K., et al. Potent and specific genetics interference by double stranded RNA in Caenorhabitis elegans. Nature, 391, 806-11 (1998).
77. Elbashir, S.M., et al., Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411, 494–498 (2001).

Figures

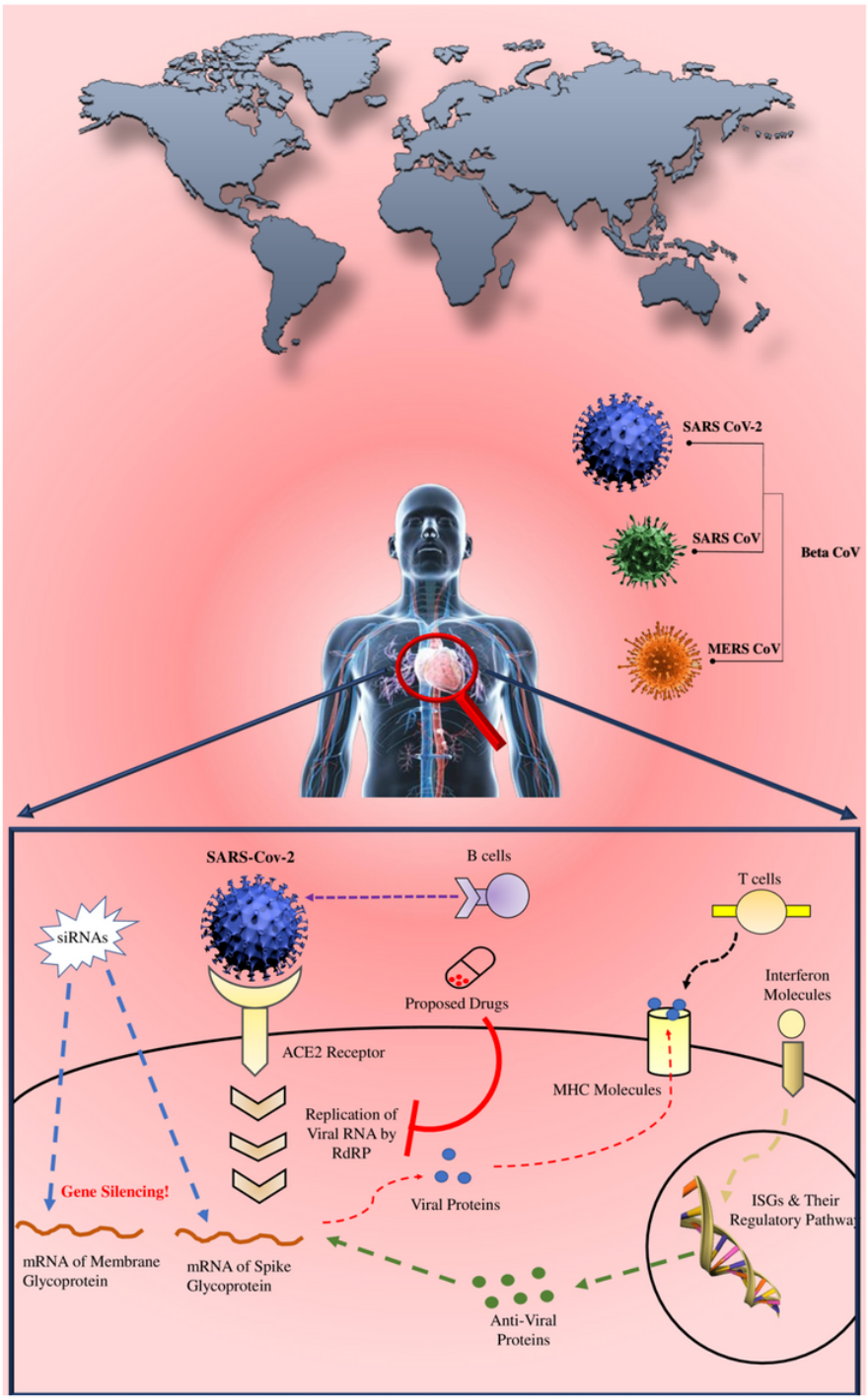

Figure 1

Overview the possible therapeutics mechanism against the pandemic-COVID-19. Here, the designed chimeric vaccine could prompt the immune response against the SARS-CoV-2 by targeting membrane and spike glycoprotein. Proposed drug compounds can inhibit the replication of viral RNA by targeting RNA dependent RNA polymerase (RdRp) protein. The designed siRNAs can silence the activity of both membrane and spike glycoproteins. Moreover, the plausible Interferon stimulating genes (ISGs) can be targeted to induce the interferons against the viral proteins of SARS-CoV-2. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square

concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

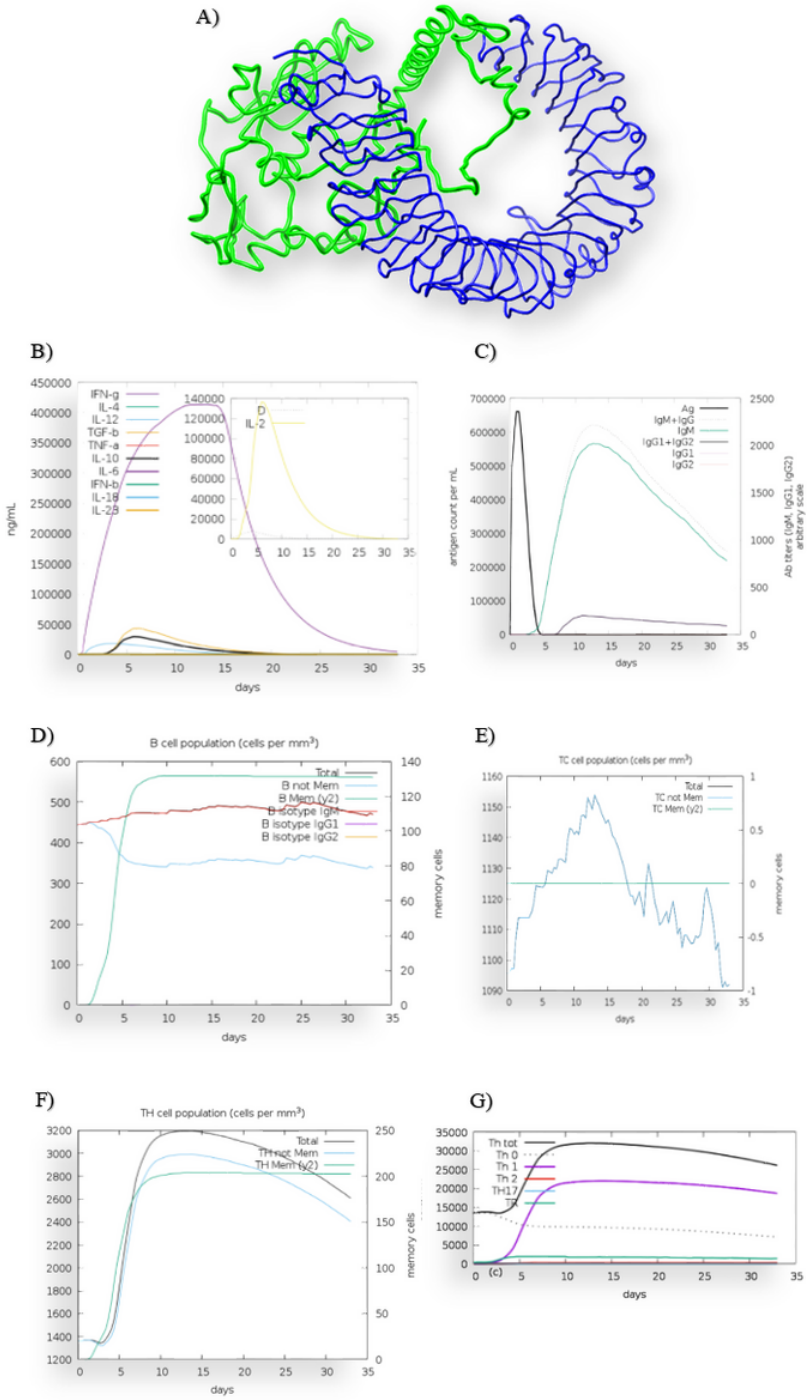

Figure 2

Innate and adaptive immune responses from the proposed vaccine. (A) The predicted interaction of the vaccine with Toll Like Receptor-4 (TLR4). Protein-Protein docking analysis demonstrated that the vaccine can interact with TLR4. The constructed vaccine was injected in the immune system simulator. (B) The interaction consequently released inflammatory and anti-inflammatory cytokines. Here, the D (Simpson index) represents the diversity of T cell population. Moreover, (C) IgM and IgG containing B cells produced with (D) generation of Memory B cell (E) Population of Cytotoxic T cell spiked and (F) Memory Helper T cell produced (G) The immune response stimulated

Th1 cells with Regulatory T cells. The immune system produced memory B cell, memory T cell, IgM and IgG. Moreover, the induction of Th1 cells and regulatory T cells were observed.

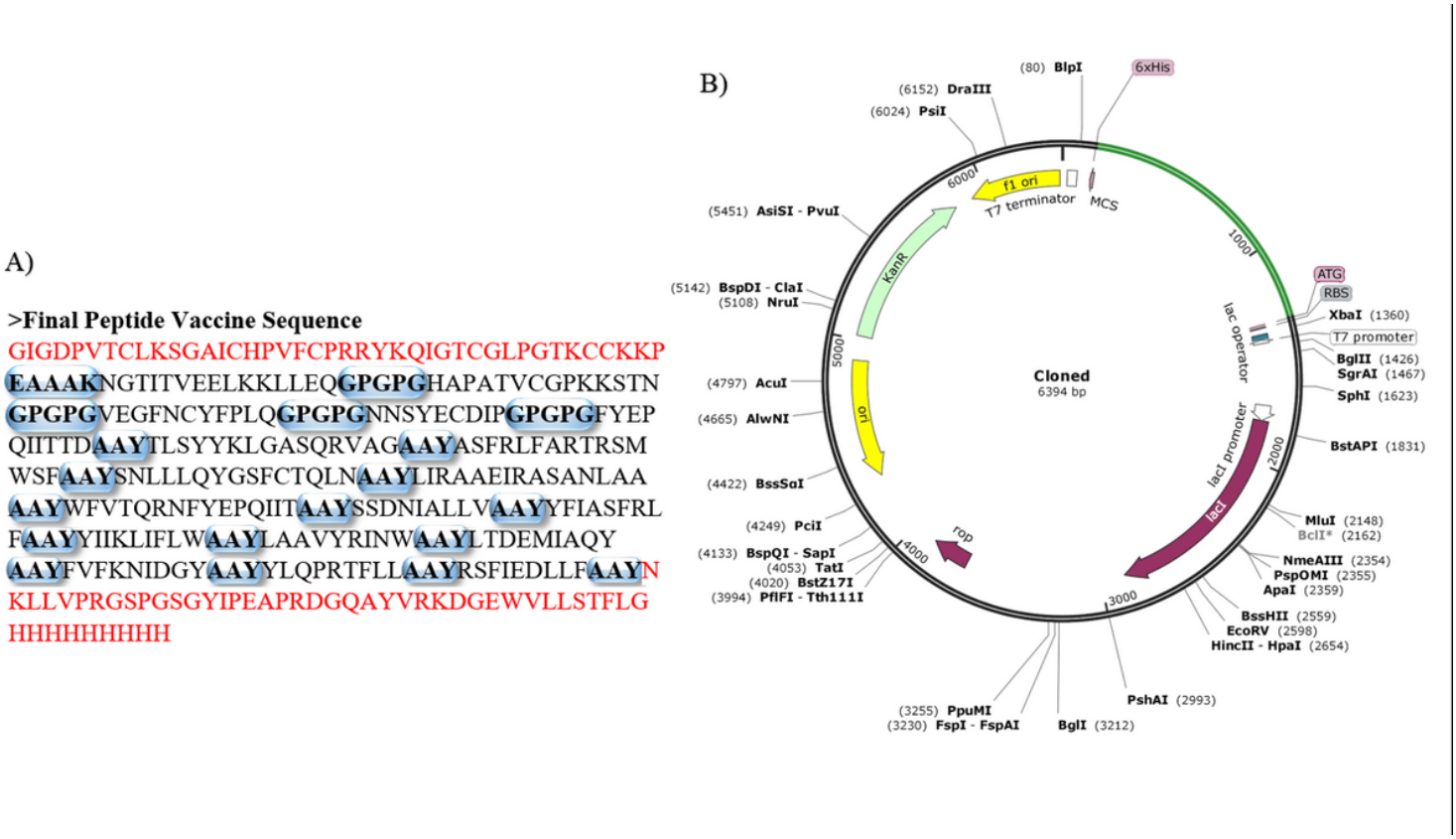

Figure 3

Production of the vaccine can be achieved through recombinant DNA technology. (A) The primary sequence of the chimeric vaccine. The red letters represent adjuvants and the bold capsules depict the peptide linkers. The adjuvants were conjugated to elicit TLR 4 mediated Th1 specific antiviral responses. The epitopes were joined with amino acid linkers. This sequence was reversely translated and (B) The DNA sequence of was inserted in pET-28a(+) expression vector for vaccine production in E.coli strain K12.

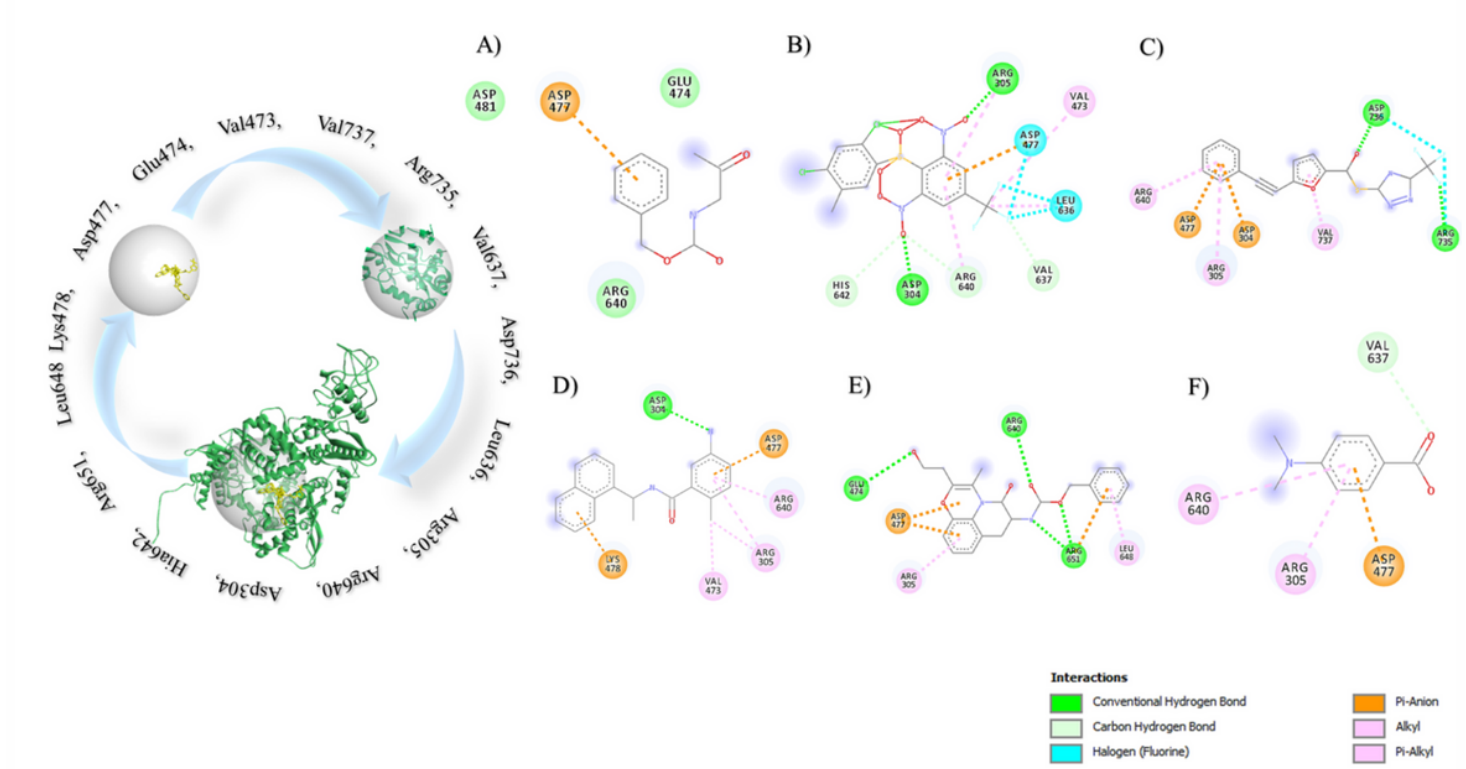

Figure 4

Drug-receptor interactions. A) benzyl (2-oxopropyl) carbamate showed the binding energy -4.2 Kcal/mol and their interacting amino acid residues are ASP481, ASP477, GLU474, ARG640; B) 2-[(2,4-dichloro-5-methylphenyl)sulfonyl]-1,3-dinitro-5-(trifluoromethyl)benzene showed the binding energy -7.6 Kcal/mol with ASP304, ARG305, VAL473, ASP477, LEU636, VAL637, ARG640, HIS642 amino acid interactions ;C) s-[5-(trifluoromethyl)-4h-1,2,4-triazol-3-yl]5-(phenylethynyl)furan-2-carbothioate docked RdRp protein with the binding energy -5.9 Kcal/mol and their interacting residues are ASP304, ARG305, ASP477, ARG735, ASP736, ARG640, VAL737; D) 5-amino-2-methyl-N-[(1R)-1-naphthalen-1-ylethyl]benzamide and RdRp protein had the docking energy about -6.3 Kcal/mol with the interacting amino acid residues ASP304, ARG305, ASP477, ARG640, VAL473, LYS478; E) Nalpha-[(benzyloxy)carbonyl]-n-[(1r)-4-hydroxy-1-methyl-2-oxobutyl]-l-phenylalaninamide had the docking enrgy-7.4 Kcal/mol and interacting residues ARG305, ASP477, GLU474, ARG640, ARG651, LEU648; F) 4-(Dimethylamino)benzoic acid also showed the significant binding energy -4.0 Kcal/mol with ARG305, ASP477, ARG640, VAL637. The most common interacting residues are Val473, Glu474, Asp477, Lys478, Leu636, Val637, Arg735, Asp736, Val737, Asp304, Arg305, Arg640, His642, Arg651 and Leu648 and all of these docked into the binding site of the RdRp protein. Here, the interactions type has been marked with different color code.

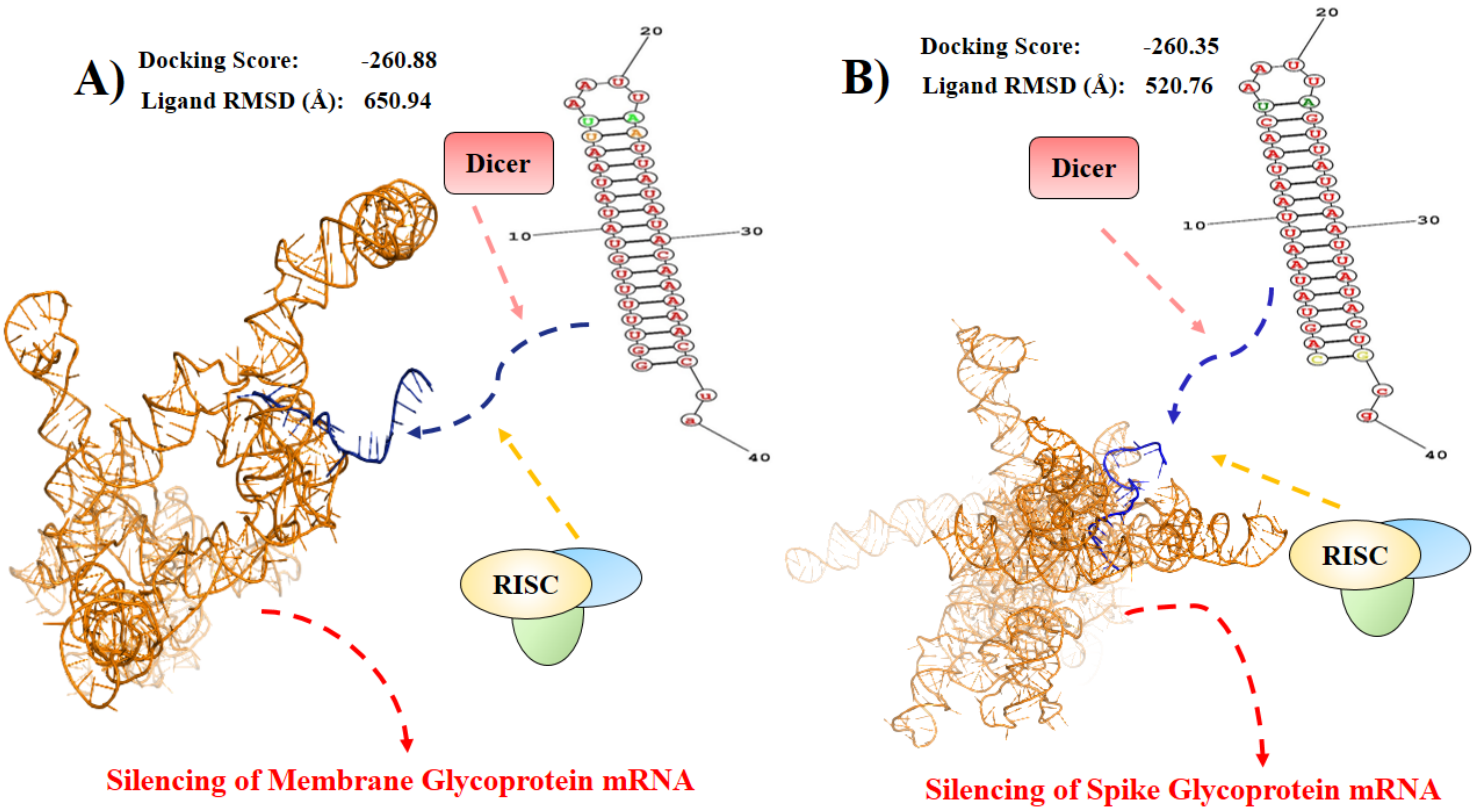

Figure 5

Predicted pathway for siRNA based viral gene silencing. In cytosolic environment the Dicer and RNA-induced silencing complex (RISC) might degrade the viral mRNAs. Here, (A) the guiding anti sense strand bonded with membrane glycoprotein mRNA (B) and spike glycoprotein mRNA. The guiding anti sense strand is very specific to the glycoproteins to silence their activity.

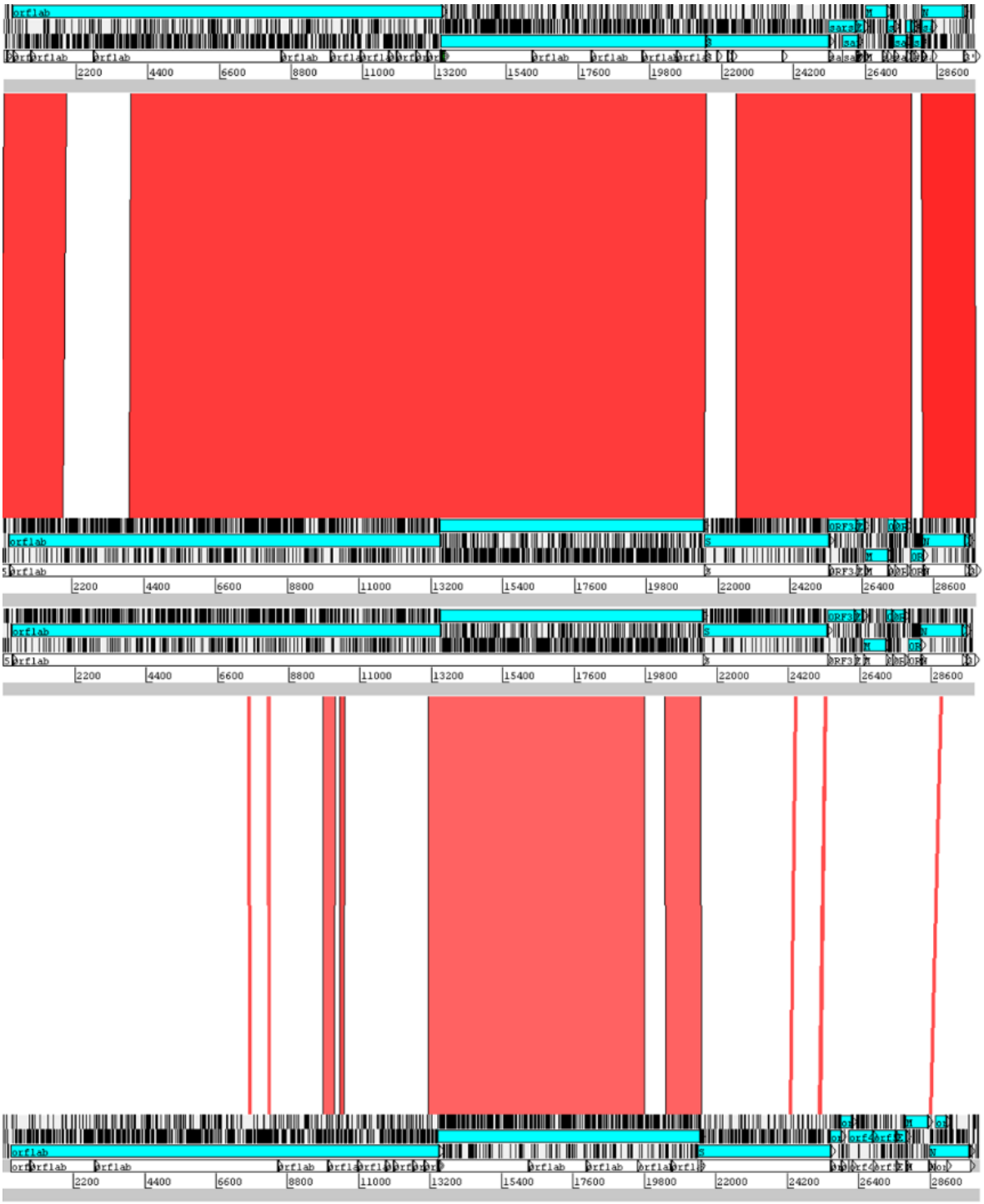

Figure 6

The whole genome comparison among theSARS-CoV, SARS-CoV-2 and MERS CoV. MERS seems to be distantly related as it showed only 29% similarity with the SARS-CoV-2. Besides, MERS CoV shared 88% similarity with the newly evolved SARS-CoV-2 virus. The variation found only in ORF1ab, Spike and ORF8 region of SARS-CoV-2 with SARS-CoV. Alternatively, orf1ab region from MERS is similar to SARS-CoV-2.

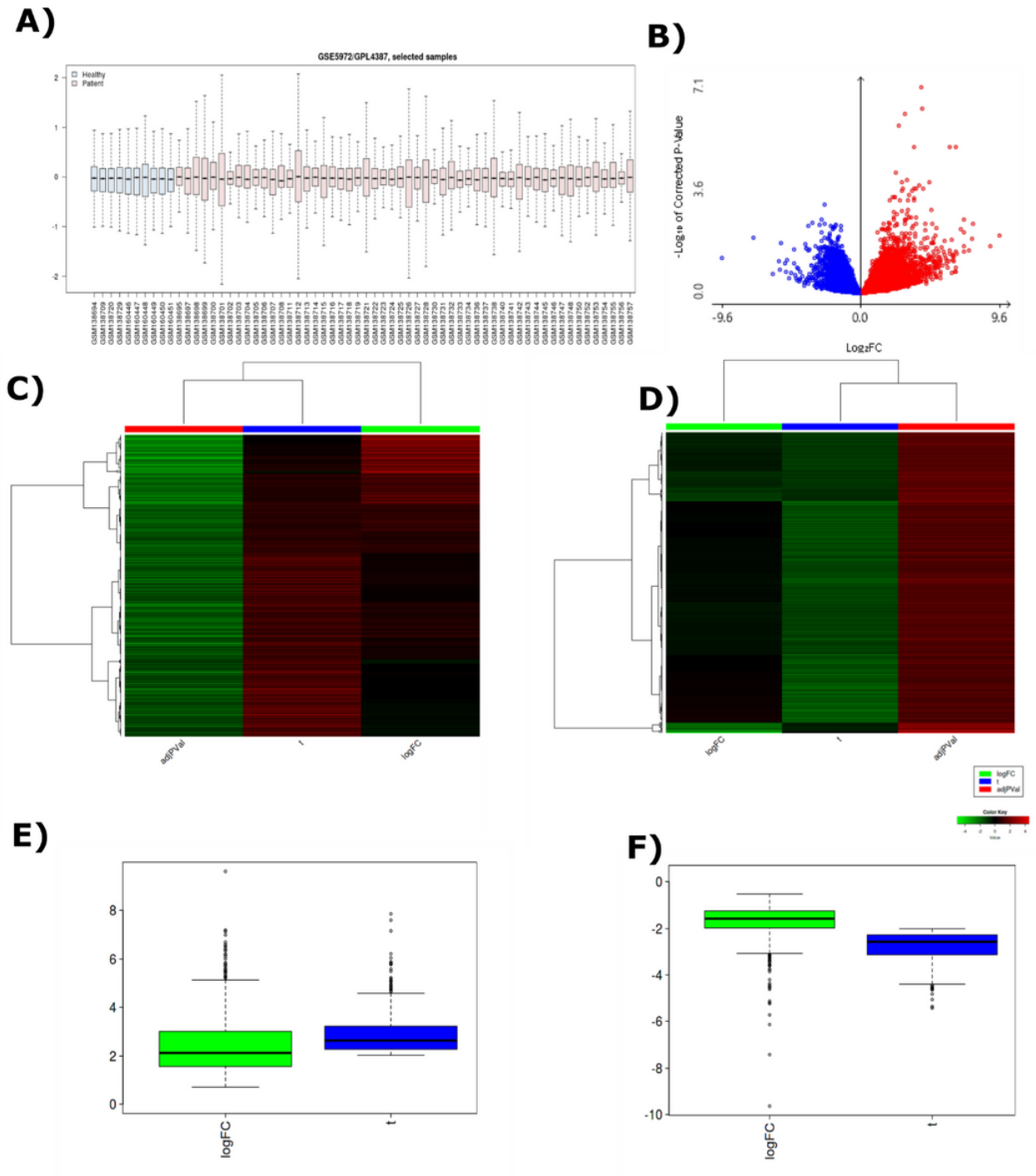

Figure 7

Analysis of Differential Gene Expression. A) Median centralized 10 healthy and 54 SARS patient samples for comparison with limma normalization method. B) LogFC against adjusted normalized p-value modelled for Gene distribution in Volcano Plot. Hierarchical Clustering Heatmap of C) Upregulated and D) Downregulated Genes. Distance was based on Correlation and Z score in where cut off was set to 4. Distance based on Correlation & Z score Cut off was set to 4. Frequency distribution of LogFC and T value of E) upregulated and F) downregulated genes.

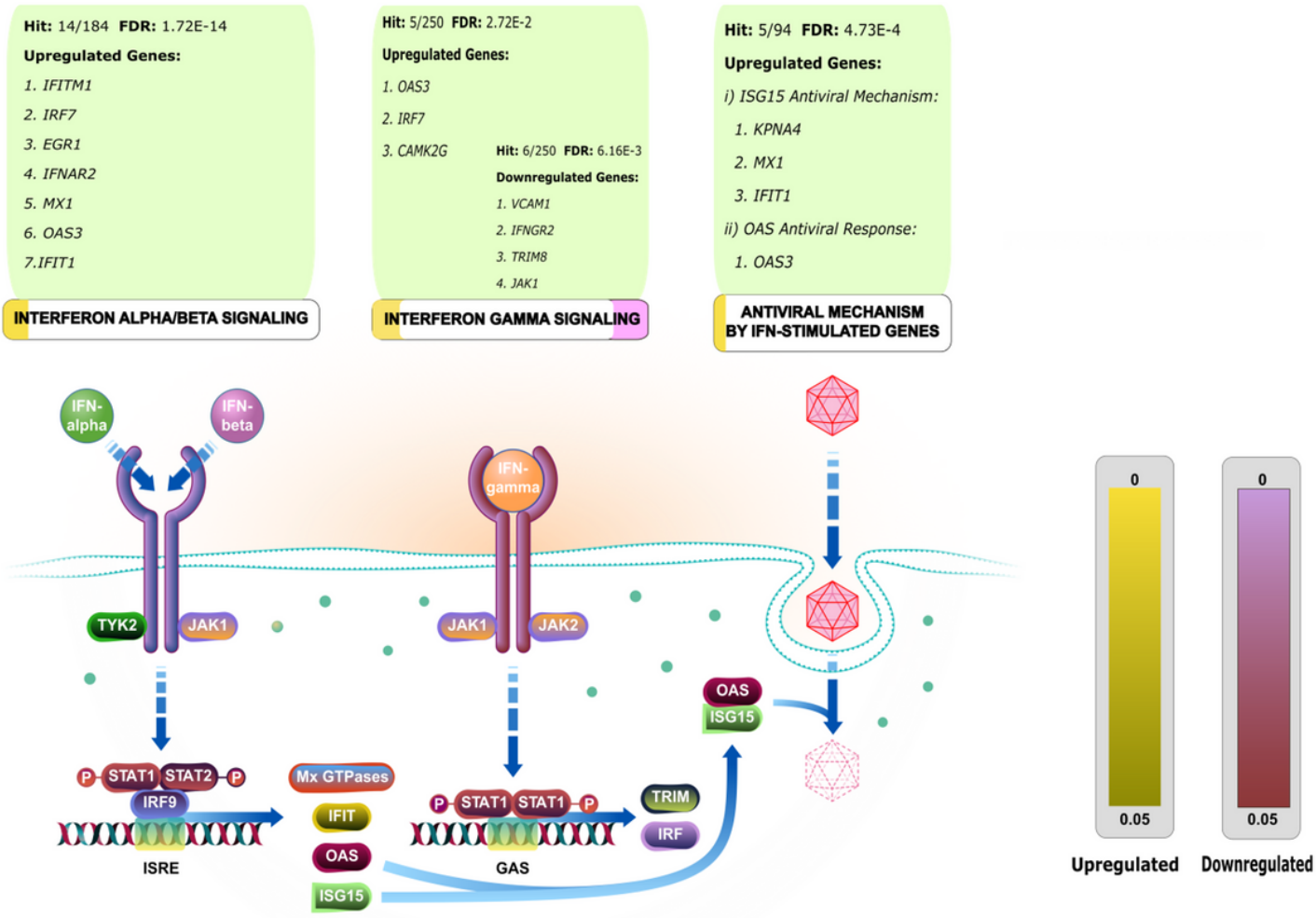

Figure 8

Interferon stimulating genes (ISGs) in pathway. Overrepresented pathway of interferon regulation by overexpressed and under expressed ISGs. The gene list including KPNA4, IRF7, OAS3, CAMK2G, IFITM1, IFIT1, EGR1, MX1 and IFNAR2 from the upregulated genes and TRIM8, VCAM1, IFNGR2 and JAK1 from the downregulated genes might stimulate the Interferon against the SARS-CoV-2. Here, the possible pathway mechanism has been explored to induce the immune response by the Interferon stimulating genes (ISGs).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryData14.Biologicalprocessofdownregulatedgenes.csv
- SupplementaryFig.2.docx
- SupplementaryFig.3.docx
- SupplementaryFig.1.docx
- SupplementaryData7.DifferentiallyExpresseddownregulatedgenesofSARSCoV.csv
- SupplementaryData4.DifferentiallyExpressedgenesofSARSCoV.csv
- SupplementaryData6.DifferentiallyExpressedUpregulatedgenesofSARSCoV.csv

- SupplementaryData1.67wholegenomesofSARSCoV2ofdifferentcountries.txt.txt
- SupplementaryTable5.docx
- SupplementaryData9.HierarchicalClusteringHeatmapofDownregulatedregulatedgenes.csv
- SupplementaryData25.InterferonlinkedISGgenesfromdownregulatedgenes.csv
- SupplementaryData24.InterferonlinkedISGgenesfromupregulatedgenes.csv
- SupplementaryData23.Interferonlinkeddownregulatedgenes.csv
- SupplementaryData22.Interferonlinkedupregulatedgenes.csv
- SupplementaryData20.InterferonstimulatinggenesISGsfromupregulatedgenes.csv
- SupplementaryData21.InterferonstimulatinggenesISGsfromdownregulatedgenes.csv
- SupplementaryData3.Consensusandconservedregionof67wholegenomesofSARSCoV2ofdifferentcountries.txt
- SupplementaryData8.HierarchicalClusteringHeatmapofUpregulatedgenes.csv
- SupplementaryTable6.docx
- SupplementaryTable1.docx
- SupplementaryData15.Molecularfunctionofdownregulatedgenes.csv
- SupplementaryData16.Cellularcomponentofdownregulatedgenes.csv
- SupplementaryData17.KEGGpathwayofdownregulatedgenes.csv
- SupplementaryData18.Viralregulatoryupregulatedgenes.csv
- SupplementaryData19.Viralregulatoryupregulatedgenes.csv
- SupplementaryData13.KEGGpathwayofupregulatedgenes.csv
- SupplementaryData10.Biologicalprocessofupregulatedgenes.csv
- SupplementaryTable2.docx
- SupplementaryFig.4.docx
- SupplementaryTable3.docx
- SupplementaryFig.6.docx
- SupplementaryData11.MolecularFunctionofupregulatedgenes.csv
- SupplementaryData2.Multiplesequencealignmentof67wholegenomesofSARSCoV2fromdifferentcountries.pdf
- SupplementaryTable4.docx
- SupplementaryFig.5.docx
- SupplementaryData5.DifferentiallyExpressedgeneswithHGNCsymbolsofSARSCoV.csv
- SupplementaryData12.Cellularcomponentofupregulatedgenes.csv