

LARGE SEQUENCE PAN-CORONAVIRUS VACCINE COMPOSITIONS

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] This application claims benefit of U.S. Provisional Application No. 63/009,907 filed April 14, 2020 and U.S. Provisional Application No. 63/084,421 filed September 28, 2020, the specification(s) of which is/are incorporated herein in their entirety by reference.

REFERENCE TO A SEQUENCE LISTING

[0002] Applicant asserts that the information recorded in the form of an Annex C/ST.25 text file submitted under Rule 13*ter*.1(a), entitled UCI_20_06B_PCT_Sequence_Listing_ST25, is identical to that forming part of the international application as filed. The content of the sequence listing is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

[0003] The present invention relates to vaccines, for example viral vaccines, such as those directed to coronaviruses, e.g., pan-coronavirus vaccines.

BACKGROUND OF THE INVENTION

[0004] Over the last two decades, there have been three deadly human outbreaks of Coronaviruses (CoVs) caused by emerging zoonotic CoVs: SARS-CoV, MERS-CoV, and the latest highly transmissible and deadly SARS-CoV-2, which has caused the current COVID-19 global pandemic. All three deadly CoVs originated from bats, the natural hosts, and transmitted to humans via various intermediate animal reservoirs (e.g., pangolins, civet cats and camels). Because there is currently no universal pan-Coronavirus vaccine available, it remains highly possible that other global COVID-like pandemics will emerge in the coming years, caused by yet another spillover of an unknown zoonotic bat-derived SARS-like Coronavirus (SL-CoV) into an unvaccinated human population.

[0005] Neutralizing antibodies and antiviral effector CD4⁺ and CD8⁺ T cells appear to be crucial in reducing viral load in the majority of infected asymptomatic and convalescent patients. However, very little information exists on the antigenic landscape and the repertoire of B-cell and CD4⁺ and CD8⁺ T cell epitopes that are conserved among human and bat Coronavirus strains.

SUMMARY OF THE INVENTION

[0006] Determining the antigen and epitope landscapes that are antigenic, immunogenic, protective and conserved among human and animal Coronaviruses as well as the repertoire, phenotype and function of B cells and CD4⁺ and CD8⁺ T cells that correlate with resistance seen in asymptomatic COVID-19 patients may inform in the development of future pan-Coronavirus vaccines. The present invention describes using several immuno-informatics and sequence alignment approaches and several immunological assays both *in vitro* in humans and *in vivo* in animal modes (e.g mice, hamster and monkeys) to identify several antigenic, immunogenic, protective highly conserved large sequences that include human B cell, CD4⁺ and CD8⁺ T cell epitopes that are highly conserved, e.g., highly conserved in: (i) greater than 81,000 SARS-CoV-2 human strains identified in 190 countries on six continents; (ii) six

circulating CoVs that caused previous human outbreaks of the "Common Cold"; (iii) nine SL-CoVs isolated from bats; (iv) nine SL-CoV isolated from pangolins; (v) three SL-CoVs isolated from civet cats; and (vi) four MERS strains isolated from camels. Furthermore, the present invention describes the identification of cross-reactive epitopes that: recalled B cell, CD4+ and CD8+ T cells from both COVID-19 patients and healthy individuals who were never exposed to SARS-CoV-2; and induced strong B cell and T cell responses in "humanized" Human Leukocyte Antigen (HLA)-DR1/HLA-A*02:01 double transgenic mice as well as in humans that do not express HLA-DR-1 or HLA-A*02:01 haplotypes. Unlike small epitopes that are restricted to certain HLA haplotype, the large sequences encompass several epitopes restricted to large numbers of HLA haplotypes, thus ascertaining large vaccine coverage of human population regardless of HLA haplotypes and regardless of race and ethnicity.

[0007] The present invention is not limited to vaccine compositions for use in humans. The present invention includes vaccine compositions for use in other pet animals such as dogs, cats, etc.

[0008] The vaccine compositions herein have the potential to provide lasting B and T cell immunity regardless of Coronaviruses mutations. This may be due at least partly because the vaccine compositions target highly conserved structural and non-structural Coronavirus antigens, such as Coronavirus nucleoprotein (also known as nucleocapsid), in combination with other Coronavirus structural and non-structural antigens with a low mutation rate found in perhaps every human and animal Coronaviruses variants and strains.

[0009] The present invention is also related to selecting highly conserved structural (e.g., spike protein) and non-structural Coronavirus antigens inside the virus (e.g., non-spike protein such as nucleocapsid), which may be viral proteins that are normally not necessarily under mutation pressure by the immune system.

[0010] The present invention provides pan-Coronavirus recombinant vaccine compositions that induces board, strong and long lasting B and T cell protective immune responses in humans and pets and animals.

[0011] In certain embodiments, the vaccine compositions are for use in humans. In certain embodiments, the vaccine compositions are for use in animals, such as but not limited to mice, cats, dogs, non-human primates, other animals susceptible to coronavirus infection, other animals that may function as preclinical animal models for coronavirus infections, etc.

[0012] As used herein, the term "multi-epitope" refers to a composition comprising more than one B and T cell epitope wherein at least: one CD4 and/or CD8 T cell epitope is MHC-restricted and recognized by a TCR, and at least one epitope is a B cell epitope. For example, the vaccine compositions herein may be multi-epitope pan-coronavirus vaccine compositions.

[0013] As used herein, the term "recombinant vaccine composition" may refer to one or more proteins or peptides encoded by one or more recombinant genes, e.g., genes that have been cloned into one or more systems that support the expression of said gene(s). The term "recombinant vaccine composition" may

refer to the recombinant genes or the system that supports the expression of said recombinant genes.

[0014] For example, the present invention provides a pan-coronavirus recombinant vaccine composition comprising one or more large sequences, wherein each of the one or more large sequences comprise at least one of: one or more conserved coronavirus B-cell target epitopes; one or more conserved coronavirus CD4+ T cell target epitopes; and/or one or more conserved coronavirus CD8+ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.

[0015] The present invention also features a pan-coronavirus recombinant vaccine composition, the composition comprising two or more large sequences, wherein each of the two or more large sequences comprise at least one of: one or more conserved coronavirus B-cell target epitopes; one or more conserved coronavirus CD4+ T cell target epitopes; and/or one or more conserved coronavirus CD8+ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.

[0016] The present invention also features a pan-coronavirus recombinant vaccine composition, the composition comprising whole spike protein; and one or both of: one or more conserved coronavirus CD4+ T cell target epitopes; and/or one or more conserved coronavirus CD8+ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.

[0017] The present invention also features a pan-coronavirus recombinant vaccine composition, the composition comprising at least a portion of spike protein, the portion of spike protein comprising a trimerized SARS-CoV-2 receptor-binding domain (RBD); and one or both of: one or more conserved coronavirus CD4+ T cell target epitopes; one or more conserved coronavirus CD8+ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.

[0018] The present invention also features a pan-coronavirus recombinant vaccine composition, the composition comprising whole spike protein; and one or more conserved coronavirus CD4+ T cell target epitopes; and one or more conserved coronavirus CD8+ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.

[0019] The present invention also features a pan-coronavirus recombinant vaccine composition, the composition comprising at least a portion of spike protein, the portion of spike protein comprising a trimerized SARS-CoV-2 receptor-binding domain (RBD); and one or more conserved coronavirus CD4+ T cell target epitopes; and one or more conserved coronavirus CD8+ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.

[0020] The present invention also features a pan-coronavirus recombinant vaccine composition, the composition comprising an antigen delivery system encoding one or more large sequences, wherein each of the one or more large sequences comprise at least one of: one or more conserved coronavirus B-cell target epitopes; one or more conserved coronavirus CD4+ T cell target epitopes; and/or one or more conserved coronavirus CD8+ T cell target epitopes; wherein at least one epitope is from a non-spike protein.

[0021] The present invention also features a pan-coronavirus recombinant vaccine composition, the composition comprising an antigen delivery system encoding two or more large sequences, wherein each of the two or more large sequences comprise at least one of: one or more conserved coronavirus B-cell target epitopes; one or more conserved coronavirus CD4+ T cell target epitopes; and/or one or more conserved coronavirus CD8+ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.

[0022] The present invention also features a pan-coronavirus recombinant vaccine composition, the composition comprising an antigen delivery system encoding whole spike protein; and one or both of: one or more conserved coronavirus CD4+ T cell target epitopes; and/or one or more conserved coronavirus CD8+ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.

[0023] The present invention also features a pan-coronavirus recombinant vaccine composition, the composition comprising an antigen delivery system encoding at least a portion of spike protein, the portion of spike protein comprising a trimerized SARS-CoV-2 receptor-binding domain (RBD); and one or both of: one or more conserved coronavirus CD4+ T cell target epitopes; one or more conserved coronavirus CD8+ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.

[0024] The present invention also features a pan-coronavirus recombinant vaccine composition, the composition comprising an antigen delivery system encoding whole spike protein; and one or more conserved coronavirus CD4+ T cell target epitopes; and one or more conserved coronavirus CD8+ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.

[0025] The present invention also features a pan-coronavirus recombinant vaccine composition, the composition comprising an antigen delivery system encoding at least a portion of spike protein, the portion of spike protein comprising a trimerized SARS-CoV-2 receptor-binding domain (RBD); and one or more conserved coronavirus CD4+ T cell target epitopes; and one or more conserved coronavirus CD8+ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.

[0026] Referring to the aforementioned compositions and the embodiments herein, in some embodiments, the the non-spike protein is ORF1ab protein, ORF3a protein, Envelope protein, Membrane glycoprotein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, Nucleocapsid protein and ORF10 protein.

[0027] In some embodiments, the one or more large sequences are highly conserved among human and animal coronaviruses. In some embodiments, the one or more large sequences are derived from at least one of SARS-CoV-2 protein. In some embodiments, the one or more large sequences are derived from one or more of: one or more SARS-CoV-2 human strains or variants in current circulation; one or more coronaviruses that has caused a previous human outbreak; one or more coronaviruses isolated from animals selected from a group consisting of bats, pangolins, civet cats, minks, camels, and other animal receptive to coronaviruses; or one or more coronaviruses that cause the common cold. In some embodiments, the one or more SARS-CoV-2 human strains or variants in current circulation are selected

from: strain B.1.177; strain B.1.160, strain B.1.1.7; strain B.1.351; strain P.1; strain B.1.427/B.1.429; strain B.1.258; strain B.1.221; strain B.1.367; strain B.1.1.277; strain B.1.1.302; strain B.1.525; strain B.1.526, strain S:677H, and strain S:677P. In some embodiments, the one or more coronaviruses that cause the common cold are selected from: 229E alpha coronavirus, NL63 alpha coronavirus, OC43 beta coronavirus, and HKU1 beta coronavirus. In some embodiments, the conserved large sequences are selected from Variants Of Concern or Variants Of Interest.

[0028] In some embodiments, the composition comprises two or more large sequences. In some embodiments, the composition comprises three or more large sequences. In some embodiments, the composition comprises two large sequences. In some embodiments, the composition comprises three large sequences. In some embodiments, the composition comprises four large sequences. In some embodiments, the composition comprises five large sequences.

[0029] In some embodiments, the large sequences are derived from structural proteins, non-structural proteins, or a combination thereof. In some embodiments, the large sequences or target epitopes are derived from a SARS-CoV-2 protein selected from a group consisting of: ORF1ab protein, Spike glycoprotein, ORF3a protein, Envelope protein, Membrane glycoprotein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, Nucleocapsid protein and ORF10 protein.

[0030] In some embodiments, the large sequence or the target epitope derived from the Spike glycoprotein is RBD. In some embodiments, the large sequence or the target epitope derived from the Spike glycoprotein is NTD. In some embodiments, the large sequence or the target epitope derived from the Spike glycoprotein includes both the RBD and NTD regions. In some embodiments, the large sequence or the target epitope derived from the spike glycoprotein are recognized by neutralizing and blocking antibodies. In some embodiments, the large sequence or the target epitope derived from the spike glycoprotein induces neutralizing and blocking antibodies. In some embodiments, the large sequence or the target epitope derived from the spike glycoprotein induces neutralizing and blocking antibodies that recognize and neutralize the virus.

[0031] In some embodiments, the large sequence or the target epitope derived from the spike glycoprotein induces neutralizing and blocking antibodies that recognize the spike protein.

[0032] In some embodiments, the ORF1ab protein comprises nonstructural protein (Nsp) 1, Nsp2, Nsp3, Nsp4, Nsp5, Nsp6, Nsp7, Nsp8, Nsp9, Nsp10, Nsp11, Nsp12, Nsp13, Nsp14, Nsp15 and Nsp16. In some embodiments, the one or more conserved coronavirus CD8+ T cell target epitopes are selected from: spike glycoprotein, Envelope protein, ORF1ab protein, ORF7a protein, ORF8a protein, ORF10 protein, or a combination thereof. In some embodiments, the one or more conserved coronavirus CD8+ T cell target epitopes are selected from: S₂₋₁₀, S₁₂₂₀₋₁₂₂₈, S₁₀₀₀₋₁₀₀₈, S₉₅₈₋₉₆₆, E₂₀₋₂₈, ORF1ab₁₆₇₅₋₁₆₈₃, ORF1ab₂₃₆₃₋₂₃₇₁, ORF1ab₃₀₁₃₋₃₀₂₁, ORF1ab₃₁₈₃₋₃₁₉₁, ORF1ab₅₄₇₀₋₅₄₇₈, ORF1ab₆₇₄₉₋₆₇₅₇, ORF7b₂₆₋₃₄, ORF8a₇₃₋₈₁, ORF1₀₃₋₁₁, and ORF1₀₅₋₁₃. In some embodiments, the one or more conserved coronavirus CD8+ T cell target epitopes are selected from SEQ ID NO: 2-29. In some embodiments, the one or more conserved coronavirus CD8+ T cell target epitopes are selected from SEQ ID NO: 30-57. In some embodiments, the one or more

conserved coronavirus CD4+ T cell target epitopes are selected from: spike glycoprotein, Envelope protein, Membrane protein, Nucleocapsid protein, ORF1a protein, ORF1ab protein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, or a combination thereof. In some embodiments, the one or more conserved coronavirus CD4+ T cell target epitopes are selected from: ORF1a₁₃₅₀₋₁₃₆₅, ORF1ab₅₀₁₉₋₅₀₃₃, ORF6₁₂₋₂₆, ORF1ab₆₀₈₈₋₆₁₀₂, ORF1ab₆₄₂₀₋₆₄₃₄, ORF1a₁₈₀₁₋₁₈₁₅, S₁₋₁₃, E₂₆₋₄₀, E₂₀₋₃₄, M₁₇₆₋₁₉₀, N₃₈₈₋₄₀₃, ORF7a₃₋₁₇, ORF7a₁₋₁₅, ORF7b₈₋₂₂, ORF7a₉₈₋₁₁₂, and ORF8₁₋₁₅. In some embodiments, the one or more conserved coronavirus CD4+ T cell target epitopes are selected from SEQ ID NO: 58-73. In some embodiments, the one or more conserved coronavirus CD4+ T cell target epitopes are selected from SEQ ID NO: 74-105. In some embodiments, the one or more conserved coronavirus B cell target epitopes are selected from Spike glycoprotein. In some embodiments, the one or more conserved coronavirus B cell target epitopes are selected from: S₂₈₇₋₃₁₇, S₅₂₄₋₅₉₈, S₆₀₁₋₆₄₀, S₈₀₂₋₈₁₉, S₈₈₈₋₉₀₉, S₃₈₉₋₃₉₃, S₄₄₀₋₅₀₁, S₁₁₃₃₋₁₁₇₂, S₃₂₉₋₃₆₃, and S₁₃₋₃₇. In some embodiments, the one or more coronavirus B cell target epitopes are selected from SEQ ID NO: 106-116. In some embodiments, the one or more coronavirus B cell target epitopes are selected from SEQ ID NO: 117-138.

[0033] In some embodiments, the one or more conserved coronavirus B cell target epitopes are in the form of a large sequence. In some embodiments, the large sequence is full length spike glycoprotein. In some embodiments, the large sequence is a partial spike glycoprotein. In some embodiments, the spike glycoprotein has two consecutive proline substitutions at amino acid positions 986 and 987. In some embodiments, the spike glycoprotein has single amino acid substitutions at amino acid positions comprising Tyr-83 and Tyr-489, Gln-24 and Asn-487. In some embodiments, the transmembrane anchor of the spike protein has an intact S1-S2 cleavage site. In some embodiments, the spike protein is in its stabilized conformation. In some embodiments, the spike protein is stabilized with proline substitutions at amino acid positions 986 and 987 at the top of the central helix in the S2 subunit.

[0034] In some embodiments, the one or more large sequences are derived from a whole protein sequence expressed by SARS-CoV-2. In some embodiments, the one or more large sequences are derived from a partial protein sequence expressed by SARS-CoV-2. In some embodiments, the one or more large conserved sequences from the spike protein is from a full-length spike glycoprotein. In some embodiments, the one or more large conserved sequences from the spike protein is from a partial spike glycoprotein. In some embodiments, the one or more large sequences comprises Spike glycoprotein (S) or a portion thereof, Nucleoprotein or a portion thereof, Membrane protein or a portion thereof, and ORF1a/b or a portion thereof. In some embodiments, the one or more large sequences comprises Spike glycoprotein (S) or a portion thereof, Nucleoprotein or a portion thereof, and ORF1a/b or a portion thereof. In some embodiments, the portion of the Spike glycoprotein is RBD. In some embodiments, the one or more large sequences is selected from the group consisting of: ORF1ab protein, Spike glycoprotein, ORF3a protein, Envelope protein, Membrane glycoprotein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, Nucleocapsid protein and ORF10 protein. In some embodiments, the ORF1ab protein comprises nonstructural protein (Nsp) 1, Nsp2, Nsp3, Nsp4, Nsp5, Nsp6, Nsp7, Nsp8, Nsp9, Nsp10, Nsp11, Nsp12, Nsp13, Nsp14, Nsp15 and Nsp16. In some embodiments, one or more of the large

sequences comprises a T-cell epitope restricted to a large number of human class 1 and class 2 HLA haplotypes and are not restricted to HLA-0201 for class 1 or HLA-DR for class 2.

[0035] In some embodiments, the large sequences are derived from structural proteins, non-structural proteins, or a combination thereof.

[0036] The present invention also features a recombinant vaccine composition comprising full-length spike protein. The present invention also features a recombinant vaccine composition comprising full-length spike protein or partial spike protein.

[0037] In some embodiments, the spike protein comprises Tyr-489 and Asn-487. In some embodiments, Tyr-489 and Asn-487 help with interaction with Tyr 83 and Gln-24 on ACE-2. In some embodiments, the spike protein comprises Gln-493. In some embodiments, Gln-493 helps with interaction with Glu-35 and Lys-31 on ACE-2. In some embodiments, the spike protein comprises Tyr-505. In some embodiments, Tyr-505 helps with interaction with Glu-37 and Arg-393 on ACE-2.

[0038] In some embodiments, the composition comprises a trimerized SARS-CoV-2 receptor-binding domain (RBD) sequence. In some embodiments, the trimerized SARS-CoV-2 receptor-binding domain (RBD) sequence is modified by the addition of a T4 fibrin-derived foldon trimerization domain. In some embodiments, the addition of a T4 fibrin-derived foldon trimerization domain increases immunogenicity by multivalent display. In some embodiments, the composition encodes the trimerized SARS-CoV-2 spike glycoprotein RBD antigen together with the one or more highly conserved structural and non-structural SARS-CoV-2 antigens. In some embodiments, the sequence for the antigen is GenBank accession number, MN908947.3. In some embodiments, the conserved large sequences are selected from the Variants Of Concern and Variants Of Interest. In some embodiments, the composition comprises a mutation 682-RRAR-685 → 682-QAAQ-685 in the S1-S2 cleavage site.

[0039] In some embodiments, the composition comprises at least one proline substitution. In some embodiments, the composition comprises at least two proline substitutions. In some embodiments, the proline substitution is at position K986 and V987. In some embodiments, the composition comprises K986P and V987P mutations.

[0040] In some embodiments, the large sequences are selected from SEQ ID NO: 182-185 (Table 1) or SEQ ID NO: 148-159 (Table 10).

[0041] In some embodiments, the composition further comprises a pharmaceutical carrier.

[0042] In some embodiments, the linker comprises T2A. In some embodiments, the linker is selected from T2A, E2A, and P2A. In some embodiments, a different linker is disposed between each open reading frame.

[0043] In some embodiments, the vaccine constructs are for humans. In some embodiments, the composition comprises human CXCL-11 and IL-7 or IL-2 or IL-15. In some embodiments, the vaccine

constructs are for animals. In some embodiments, the composition comprises animal CXCL-11 and IL-7 or IL-2 or IL-15. In some embodiments, the animals are cats and dogs.

[0044] In some embodiments, the delivery system is an adenovirus system. In some embodiments, the adenovirus delivery system is Ad26, Ad5, Ad35, or a combination thereof. In some embodiments, one or more of the large sequences are operatively linked to a generic promoter. In some embodiments, the generic promoter is a CMV or a CAG promoter. In some embodiments, the one or more large sequences are operatively linked to a lung-specific promoter. In some embodiments, the lung-specific promoter is SpB or CD144. In some embodiments, the composition further comprises a T cell attracting chemokine.

[0045] In some embodiments, the antigen delivery system further encodes a T cell attracting chemokine. In some embodiments, the antigen delivery system comprises two delivery systems, wherein a second delivery system encodes the T cell attracting chemokine. In some embodiments, the T cell attracting chemokine is CCL5, CXCL9, CXCL10, CXCL11, or a combination thereof. In some embodiments, the T cell attracting chemokine is operatively linked to a lung-specific promoter. In some embodiments, the T cell attracting chemokine is operatively linked to a generic promoter. In some embodiments, the composition further comprises a composition that promotes T cell proliferation.

[0046] In some embodiments, the antigen delivery system further encodes a composition that promotes T cell proliferation. In some embodiments, the antigen delivery system comprises two delivery systems, wherein a second delivery system encodes the composition that promotes T cell proliferation. In some embodiments, the composition that promotes T cell proliferation is IL-7, IL-2, or IL-15. In some embodiments, the composition that promotes T cell proliferation is operatively linked to a lung-specific promoter. In some embodiments, the composition that promotes T cell proliferation is operatively linked to a generic promoter. In some embodiments, the T cell attracting chemokine and the composition that promotes T cell proliferation are driven by the same promoter. In some embodiments, the vaccine further encodes a peptide comprising a T cell attracting chemokine and a composition that promotes T cell proliferation. In some embodiments, the peptide is operatively linked to a lung-specific promoter. In some embodiments, the peptide is operatively linked to a generic promoter. In some embodiments, the lung-specific promoter is SpB or CD144. In some embodiments, the generic promoter is a CMV or a CAG promoter.

[0047] In some embodiments, the antigen delivery system further encodes a molecular adjuvant. In some embodiments, the antigen delivery system comprises two delivery systems, wherein a second delivery system encodes the molecular adjuvant. In some embodiments, the molecular adjuvant is CpG. In some embodiments, the molecular adjuvant is a CpG polymer. In some embodiments, the molecular adjuvant is flagellin. In some embodiments, the molecular adjuvant is operatively linked to a promoter. In some embodiments, the promoter is a lung-specific promoter or a generic promoter.

[0048] In some embodiments, one or more of the large sequences are separated by a linker. In some embodiments, each of the large sequences are separated by a linker. In some embodiments, the linker is from 2 to 10 amino acids in length.

[0049] In some embodiments, the recombinant vaccine composition comprises a tag, e.g., one or more of the large sequence comprises a tag. In some embodiments, the tag is a His tag.

[0050] The present invention also includes a rVSV-panCoV recombinant vaccine composition comprising any of the vaccine compositions herein.

[0051] The present invention also includes a rAdV-panCoV recombinant vaccine composition comprising any of the vaccine compositions herein.

[0052] In some embodiments, the compositions are for use as a vaccine. In some embodiments, the compositions are for use as immunotherapy for the prevention and treatment of Coronaviruses infections and diseases. In some embodiments, the composition is used to prevent a coronavirus disease in a subject. In some embodiments, the composition is used to prevent a coronavirus infection prophylactically in a subject. In some embodiments, the composition elicits an immune response in a subject. In some embodiments, the composition prolongs an immune response induced by the pan-coronavirus recombinant vaccine composition and increases T-cell migration to the lungs.

[0053] The present invention also includes a pan-coronavirus recombinant vaccine composition comprising SEQ ID NO: 139-147 (Table 10).

[0054] Non-spike proteins include any of the coronavirus proteins other than spike, such as but not limited to Envelope protein, Membrane protein, Nucleocapsid protein, ORF1a protein, ORF1ab protein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, etc.

[0055] In certain embodiments, the compositions of the present invention, e.g., the large sequences, comprise one or more conserved target epitopes, e.g., one or more conserved coronavirus B-cell target epitopes; one or more conserved coronavirus CD4+ T cell target epitopes; and/or one or more conserved coronavirus CD8+ T cell target epitopes. In some embodiments, a conserved target epitope is one that is one of the 5 most conserved epitopes (for its epitope type, e.g., B cell, CD4 T cell, CD8 T cell) identified in a sequence alignment and analysis. In some embodiments, a conserved target epitope is one that is one of the 10 most conserved epitopes (for its epitope type, e.g., B cell, CD4 T cell, CD8 T cell) identified in a sequence alignment and analysis. In some embodiments, a conserved target epitope is one that is one of the 15 most conserved epitopes (for its epitope type, e.g., B cell, CD4 T cell, CD8 T cell) identified in a sequence alignment and analysis. In some embodiments, a conserved target epitope is one that is one of the 20 most conserved epitopes (for its epitope type, e.g., B cell, CD4 T cell, CD8 T cell) identified in a sequence alignment and analysis. In some embodiments, a conserved target epitope is one that is one of the 25 most conserved epitopes (for its epitope type, e.g., B cell, CD4 T cell, CD8 T cell) identified in a sequence alignment and analysis. In some embodiments, a conserved target epitope is one that is one of the 30 most conserved epitopes (for its epitope type, e.g., B cell, CD4 T cell, CD8 T cell) identified in a sequence alignment and analysis. In some embodiments, a conserved target epitope is one that is one of the 35 most conserved epitopes (for its epitope type, e.g., B cell, CD4 T cell, CD8 T cell) identified in a sequence alignment and analysis. In some embodiments, a conserved target epitope is one that is one of

the 40 most conserved epitopes (for its epitope type, e.g., B cell, CD4 T cell, CD8 T cell) identified in a sequence alignment and analysis. In some embodiments, a conserved target epitope is one that is one of the 50 most conserved epitopes (for its epitope type, e.g., B cell, CD4 T cell, CD8 T cell) identified in a sequence alignment and analysis. Examples of sequence alignments and analyses. Are described herein. For example, steps or methods for selecting or identifying conserved large sequences may first include performing a sequence alignment and analysis of a particular number of coronavirus sequences to determine sequence similarity or identity amongst the group of analyzed sequences. In some embodiments, the sequences used for alignments may include human and animal sequences. In certain embodiments, the sequences used for alignments include one or more SARS-CoV-2 human strains or variants in current circulation; one or more coronaviruses that has caused a previous human outbreak; one or more coronaviruses isolated from animals selected from a group consisting of bats, pangolins, civet cats, minks, camels, and other animal receptive to coronaviruses; and/or one or more coronaviruses that cause the common cold. In some embodiments, the conserved large sequences are identified by: performing a sequence alignment and analysis of a particular number of coronavirus sequences to determine sequence similarity or identity amongst the group of analyzed sequences. The conserved large sequences are those that are among the most highly conserved sequences identified in the analysis. For example, the conserved large sequences may be the 2 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 5 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 8 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 10 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 15 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 20 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 30 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 40 most highly conserved sequences identified. The present invention is not limited to the aforementioned thresholds. In some embodiments, the alignment and analysis for 50 or more sequences, 100 or more sequences, 200 or more sequences, 300 or more sequences, 400 or more sequences, 500 or more sequences, 1000 or more sequences, 2000 or more sequences, 3000 or more sequences, 4000 or more sequences, 5000 or more sequences, 10,000 or more sequences, 15,00 or more sequences, more than 15,000 sequences, etc., In some embodiments, the sequences used for alignments may include human and animal sequences. In certain embodiments, the sequences used for alignments include one or more SARS-CoV-2 human strains or variants in current circulation; one or more coronaviruses that has caused a previous human outbreak; one or more coronaviruses isolated from animals selected from a group consisting of bats, pangolins, civet cats, minks, camels, and other animal receptive to coronaviruses; and/or one or more coronaviruses that cause the common cold. In some embodiments, the one or more SARS-CoV-2 human strains or variants in current circulation are selected from: strain B.1.177; strain B.1.160, strain B.1.1.7; strain B.1.351; strain P.1; strain B.1.427/B.1.429; strain B.1.258; strain B.1.221; strain B.1.367; strain B.1.1.277; strain B.1.1.302; strain B.1.525; strain B.1.526, strain S:677H, and strain S:677P. In some embodiments, the one or more coronaviruses that cause the common cold are selected from: 229E alpha coronavirus, NL63 alpha

coronavirus, OC43 beta coronavirus, and HKU1 beta coronavirus. As discussed herein, the one or more conserved large sequences comprising target epitopes, are highly conserved among human and animal coronaviruses. For any of the embodiments herein, the epitopes that are selected may be those that achieve a particular score in a binding assay (for binding to an HLA molecule, for example.)

[0056] In certain embodiments, the one or more conserved coronavirus CD8+ T cell target epitopes are selected from: spike glycoprotein, Envelope protein, ORF1ab protein, ORF7a protein, ORF8a protein, ORF10 protein, or a combination thereof. In certain embodiments, the one or more conserved coronavirus CD8+ T cell target epitopes are selected from: S₂₋₁₀, S₁₂₂₀₋₁₂₂₈, S₁₀₀₀₋₁₀₀₈, S₅₅₈₋₅₆₆, E₂₀₋₂₈, ORF1ab₁₆₇₅₋₁₆₈₃, ORF1ab₂₃₆₃₋₂₃₇₁, ORF1ab₃₀₁₃₋₃₀₂₁, ORF1ab₃₁₈₃₋₃₁₉₁, ORF1ab₅₄₇₀₋₅₄₇₈, ORF1ab₆₇₄₉₋₆₇₅₇, ORF7b₂₆₋₃₄, ORF8a₇₃₋₈₁, ORF1₀₃₋₁₁, and ORF1₀₅₋₁₃. In certain embodiments, the one or more conserved coronavirus CD8+ T cell target epitopes are selected from SEQ ID NO: 2-29. In certain embodiments, the one or more conserved coronavirus CD8+ T cell target epitopes are selected from SEQ ID NO: 30-57.

[0057] In certain embodiments, the one or more conserved coronavirus CD4+ T cell target epitopes are selected from: spike glycoprotein, Envelope protein, Membrane protein, Nucleocapsid protein, ORF1a protein, ORF1ab protein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, or a combination thereof. In certain embodiments, the one or more conserved coronavirus CD4+ T cell target epitopes are selected from: ORF1a₁₃₅₀₋₁₃₆₅, ORF1ab₅₀₁₉₋₅₀₃₃, ORF6₁₂₋₂₆, ORF1ab₆₀₈₈₋₆₁₀₂, ORF1ab₆₄₂₀₋₆₄₃₄, ORF1a₁₈₀₁₋₁₈₁₅, S₁₋₁₃, E₂₆₋₄₀, E₂₀₋₃₄, M₁₇₆₋₁₉₀, N₃₈₈₋₄₀₃, ORF7a₃₋₁₇, ORF7a₁₋₁₅, ORF7b₈₋₂₂, ORF7a₉₈₋₁₁₂, and ORF8₁₋₁₅. In certain embodiments, the one or more conserved coronavirus CD4+ T cell target epitopes are selected from SEQ ID NO: 58-73. In certain embodiments, the one or more conserved coronavirus CD4+ T cell target epitopes are selected from SEQ ID NO: 74-105.

[0058] In certain embodiments, the one or more conserved coronavirus B cell target epitopes are selected from Spike glycoprotein. In certain embodiments, the one or more conserved coronavirus B cell target epitopes are selected from: S₂₈₇₋₃₁₇, S₅₂₄₋₅₉₈, S₆₀₁₋₆₄₀, S₈₀₂₋₈₁₉, S₈₈₈₋₉₀₉, S₃₆₉₋₃₉₃, S₄₄₀₋₅₀₁, S₁₁₃₃₋₁₁₇₂, S₃₂₉₋₃₆₃, and S₁₃₋₃₇. In certain embodiments, the one or more coronavirus B cell target epitopes are selected from SEQ ID NO: 106-116. In certain embodiments, the one or more coronavirus B cell target epitopes are selected from SEQ ID NO: 117-138.

[0059] As previously discussed, in certain embodiments, the one or more conserved coronavirus B cell target epitopes are in the form of a large sequence, e.g., whole spike protein or partial spike protein (e.g., a portion of whole spike protein). In some embodiments, the whole spike protein or portion thereof is in its stabilized conformation. In certain embodiments, the transmembrane anchor of the spike protein (or portion thereof) has an intact S1–S2 cleavage site. In certain embodiments, the spike glycoprotein has two consecutive proline substitutions at amino acid positions 986 and 987, e.g., for stabilization. In certain embodiments, the spike protein or portion thereof has an amino acid substitution at amino acid position Tyr-83. In certain embodiments, the spike protein or portion thereof has an amino acid substitution at amino acid position Tyr-489. In certain embodiments, the spike protein or portion thereof has an amino acid substitution at amino acid position Gln-24. In certain embodiments, the spike protein or portion

thereof has an amino acid substitution at amino acid position Asn-487. In certain embodiments, the spike protein or portion thereof has an amino acid substitution at one or more of: Tyr-83, Tyr-489, Gln-24, Gln-493, and Asn-487, e.g., the spike protein or portion thereof may comprise Tyr-489 and Asn-487, the spike protein or portion thereof may comprise Gln-493, the spike protein or portion thereof may comprise Tyr-505, etc. Tyr-489 and Asn-487 may help with interaction with Tyr 83 and Gln-24 on ACE-2. Gln-493 may help with interaction with Glu-35 and Lys-31 on ACE-2. Tyr-505 may help with interaction with Glu-37 and Arg-393 on ACE-2.

[0060] In certain embodiments, the composition comprises a mutation 682-RRAR-685 → 682-QQAQ-685 in the S1-S2 cleavage site. In certain embodiments, the composition comprises at least one proline substitution. In certain embodiments, the composition comprises at least two proline substitutions, e.g., at position K986 and V987.

[0061] In certain embodiments, a large sequence derived from the spike glycoprotein is RBD. In certain embodiments, a large sequence derived from the spike glycoprotein is NTD. In certain embodiments, a large sequence derived from the spike glycoprotein is one or more large sequences, e.g., comprising both the RBD and NTD regions. In certain embodiments, a large sequence derived from the spike glycoprotein is recognized by neutralizing and blocking antibodies. In certain embodiments, a large sequence derived from the spike glycoprotein induces neutralizing and blocking antibodies. In certain embodiments, a large sequence derived from the spike glycoprotein induces neutralizing and blocking antibodies that recognize and neutralize the virus. In certain embodiments, a large sequence derived from the spike glycoprotein induces neutralizing and blocking antibodies that recognize the spike protein.

[0062] In certain embodiments, linkers are used, e.g., between epitopes, between large sequences, etc. In certain embodiments, the linker is from 2-10 amino acids in length. In certain embodiments, the linker is from 3-12 amino acids in length. In certain embodiments, the linker is from 5-15 amino acids in length. In certain embodiments, the linker is 10 or more amino acids in length. Non-limiting examples of linkers include AAY, KK, and GPGPG (SEQ ID NO: 186).

[0063] In some embodiments, the composition comprises the addition of a T4 fibrin-derived foldon trimerization domain. In some embodiments, the addition of a T4 fibrin-derived foldon trimerization domain increases immunogenicity by multivalent display.

[0064] In certain embodiments, the composition further comprises a T cell attracting chemokine. For example, the composition may further comprise one or a combination of CCL5, CXCL9, CXCL10, CXCL11, or a combination thereof.

[0065] In certain embodiments, the composition further comprises a composition that promotes T cell proliferation. For example, the composition may further comprise IL-7, IL-15, IL-2, or a combination thereof.

[0066] In certain embodiments, the composition further comprises a molecular adjuvant. For example, the composition may further comprise one or a combination of CpG (e.g., CpG polymer) or flagellin.

[0067] In certain embodiments, the composition comprises a tag. For example, one or more of the large sequences may comprise a tag. In certain embodiments, the epitopes are in the form of two or more antigens, wherein one or more of the antigens comprise a tag. Non-limiting examples of tags include a His tag.

[0068] In certain embodiments, the "antigen delivery system" may refer to two delivery systems, e.g., a portion of the large sequences (or other components such as chemokines, etc.) may be encoded by one delivery system and a portion of the large sequences (or other components) may be encoded by a second delivery system (or a third delivery system, etc.).

[0069] Referring to the antigen delivery system, in certain embodiments the antigen delivery system is a vesicular stomatitis virus (VSV) vector. In certain embodiments, the antigen delivery system is an adenovirus (e.g., Ad26, Ad5, Ad35, etc.)

[0070] The large sequences are operatively linked to a promoter. In certain embodiments, the promoter is a generic promoter (e.g., CMV, CAG, etc.). In certain embodiments, the promoter is a lung-specific promoter (e.g., SpB, CD144). In certain embodiments, large sequences are operatively linked to the same promoter. In certain embodiments, one or more of the large sequences are operatively linked to a first promoter and one or more large sequences are operatively linked to a second promoter. In certain embodiments, the large sequences are operatively linked to two or more promoters, e.g., a portion are operatively linked to a first promoter, a portion are operatively linked to a second promoter, etc. In certain embodiments, the large sequences are operatively linked to three or more promoters, e.g., a portion is operatively linked to a first promoter, a portion is operatively linked to a second promoter, a portion is operatively linked to a third promoter, etc. In certain embodiments, the first promoter is the same as the second promoter. In certain embodiments the second promoter is different from the first promoter. In certain embodiments, the promoter is a generic promoter (e.g., CMV, CAG, etc.). In certain embodiments, the promoter is a lung-specific promoter (e.g., SpB, CD144) promoter.

[0071] In certain embodiments, the antigen delivery system or a separate antigen delivery system encodes a T cell attracting chemokine. In certain embodiments, the antigen delivery system or a separate antigen delivery system encodes a composition that promotes T cell proliferation. In certain embodiments, the antigen delivery system or a separate antigen delivery system encodes both a T cell attracting chemokine and a composition that promotes T cell proliferation. In certain embodiments, the antigen delivery system or a separate antigen delivery system encodes a molecular adjuvant. In certain embodiments, the antigen delivery system or a separate antigen delivery system encodes a T cell attracting chemokine, a composition that promotes T cell proliferation and a molecular adjuvant. In certain embodiments, the antigen delivery system or a separate antigen delivery system encodes a T cell attracting chemokine and a molecular adjuvant. In some embodiments, the antigen delivery system or a separate antigen delivery system encodes a composition that promotes T cell proliferation and a molecular adjuvant.

[0072] In certain embodiments, the T cell attracting chemokine is CCL5, CXCL9, CXCL10, CXCL11, or a

combination thereof. In certain embodiments, the composition that promotes T cell proliferation is IL-7 or IL-15 or IL-2. In some embodiments, the molecular adjuvant is CpG (e.g., CpG polymer), flagellin, etc.).

[0073] In certain embodiments, the T cell attracting chemokine is operatively linked to a lung-specific promoter (e.g., SpB, CD144). In certain embodiments, the T cell attracting chemokine is operatively linked to a generic promoter (e.g., CMV, CAG, etc.). In certain embodiments, the composition that promotes T cell proliferation is operatively linked to a lung-specific promoter (e.g., SpB, CD144). In certain embodiments, the composition that promotes T cell proliferation is operatively linked to a generic promoter (e.g., CMV, CAG, etc.). In certain embodiments, the molecular adjuvant is operatively linked to a lung-specific promoter (e.g., SpB, CD144). In certain embodiments, the molecular adjuvant is operatively linked to a generic promoter (e.g., CMV, CAG, etc.). In certain embodiments, the T cell attracting chemokine and the composition that promotes T cell proliferation are driven by the same promoter. In certain embodiments, the T cell attracting chemokine and the composition that promotes T cell proliferation are driven by different promoters. In certain embodiments, the molecular adjuvant, the T cell attracting chemokine, and the composition that promotes T cell proliferation are driven by the same promoter. In certain embodiments, the molecular adjuvant, the T cell attracting chemokine, and the composition that promotes T cell proliferation are driven by different promoters. In certain embodiments, the molecular adjuvant and the composition that promotes T cell proliferation are driven by different promoters. In certain embodiments, the molecular adjuvant and the T cell attracting chemokine are driven by different promoters.

[0074] In certain embodiments, the T cell attracting chemokine and the composition promoting T cell proliferation are separated by a linker. In certain embodiments, the linker comprises T2A. In certain embodiments, the linker comprises E2A. In certain embodiments, the linker comprises P2A. In certain embodiments, the linker is selected from T2A, E2A, and P2A.

[0075] Referring to the antigen delivery system, in certain embodiments, a linker is disposed between each open reading frame. In certain embodiments, a different linker is disposed between each open reading frame. In certain embodiments, the same linker may be used between particular open reading frames and a different linker may be used between other open reading frames.

[0076] In some embodiments, the vaccine composition is administered using an adenovirus.

[0077] The composition herein may be used to prevent a coronavirus disease in a subject. The composition herein may be used to prevent a coronavirus infection prophylactically in a subject. The composition herein may be used to elicit an immune response in a subject. The term "subject" herein may refer to a human, a non-human primate, an animal such as a mouse, rat, cat, dog, other animal that is susceptible to coronavirus infection, or other animal used for preclinical modeling. The composition herein may prolong an immune response induced by the pan-coronavirus recombinant vaccine composition and increases T-cell migration to the lungs. In certain embodiments, the composition induces resident memory T cells (Trm). In some embodiments, the vaccine composition induces efficient and powerful protection against the coronavirus disease or infection. In some embodiments, the vaccine composition induces

production of antibodies (Abs), CD4+ T helper (Th1) cells, and CD8+ cytotoxic T-cells (CTL). In some embodiments, the composition that promotes T cell proliferation helps to promote long term immunity. In some embodiments, the T-cell attracting chemokine helps pull T-cells from circulation into the lungs.

[0078] In certain embodiments, the composition further comprises a pharmaceutical carrier.

[0079] The present invention includes any of the vaccine compositions described herein, e.g., the aforementioned vaccine compositions for delivery with nanoparticles, e.g., lipid nanoparticles. For example, the present invention includes the vaccine compositions herein encapsulated in a lipid nanoparticle.

[0080] The present invention includes the compositions described herein comprising and/or encoding a trimerized SARS-CoV-2 receptor-binding domain (RBD) and one or more highly conserved SARS-CoV-2 sequences selected from structural proteins (e.g., nucleoprotein, etc.) and non-structural protein (e.g., Nsp4, etc.). In some embodiments, the trimerized SARS-CoV-2 receptor-binding domain (RBD) sequence is modified by the addition of a T4 fibrin-derived foldon trimerization domain. In some embodiments, the addition of a T4 fibrin-derived foldon trimerization domain increases immunogenicity by multivalent display.

[0081] The present invention also features methods of producing apan-coronavirus recombinant vaccine compositions of the present invention.

[0082] For example, in some embodiments, the method comprises selecting at least conserved large sequences comprising: one or more coronavirus B-cell epitopes; one or more coronavirus CD4+ T cell epitopes; one or more coronavirus CD8+ T cell epitopes. In other embodiments, the method comprises selecting at least two conserved large sequences comprising: one or more coronavirus B-cell epitopes; one or more coronavirus CD4+ T cell epitopes; one or more coronavirus CD8+ T cell epitopes. At least one large sequence is derived from a non-spike protein. The method further comprises synthesizing an antigen or antigens comprising the selected large sequences. In some embodiments, the method comprises selecting: one or more conserved large sequences comprising one or more coronavirus B-cell epitopes; one or more coronavirus CD4+ T cell epitopes; and one or more coronavirus CD8+ T cell epitopes. At least one large sequence is derived from a non-spike protein. The method further comprises synthesizing an antigen or antigens comprising the selected large sequences. In some embodiments, the method further comprises introducing the vaccine composition to a pharmaceutical carrier. The steps for selecting the one or more conserved large sequences are disclosed herein. Methods for synthesizing recombinant proteins are well known to one of ordinary skill in the art. The vaccine compositions are disclosed herein. In some embodiments, the vaccine composition is in the form of DNA, RNA, modified RNA, protein (or peptide), or a combination thereof.

[0083] In some embodiments, the method comprises selecting: at least one conserved large sequence comprising: one or more coronavirus B-cell epitopes; one or more coronavirus CD4+ T cell epitopes; and one or more coronavirus CD8+ T cell epitopes. At least one large sequence is derived from a non-spike

protein. The method further comprises synthesizing an antigen delivery system encoding the selected large sequences. In some embodiments, the method further comprises introducing the vaccine composition to a pharmaceutical carrier. The steps for selecting the one or more conserved large sequences are disclosed herein. Methods for synthesizing antigen delivery systems are well known to one of ordinary skill in the art. The vaccine compositions are disclosed herein. In some embodiments, the vaccine composition is in the form of DNA, RNA, modified RNA, protein (or peptide), or a combination thereof.

[0084] As an example, steps or methods for selecting or identifying conserved large sequences may first include performing a sequence alignment and analysis of a particular number of coronavirus sequences, e.g., 50 or more sequences, 100 or more sequences, 200 or more sequences, 300 or more sequences, 400 or more sequences, 500 or more sequences, 1000 or more sequences, 2000 or more sequences, 3000 or more sequences, 4000 or more sequences, 5000 or more sequences, 10,000 or more sequences, 15,000 or more sequences, more than 15,000 sequences, etc., to determine sequence similarity or identity amongst the group of analyzed sequences. In some embodiments, the sequences used for alignments may include human and animal sequences. In certain embodiments, the sequences used for alignments include one or more SARS-CoV-2 human strains or variants in current circulation; one or more coronaviruses that has caused a previous human outbreak; one or more coronaviruses isolated from animals selected from a group consisting of bats, pangolins, civet cats, minks, camels, and other animal receptive to coronaviruses; and/or one or more coronaviruses that cause the common cold. In some embodiments, the one or more SARS-CoV-2 human strains or variants in current circulation are selected from: strain B.1.177; strain B.1.160, strain B.1.1.7; strain B.1.351; strain P.1; strain B.1.427/B.1.429; strain B.1.258; strain B.1.221; strain B.1.367; strain B.1.1.277; strain B.1.1.302; strain B.1.525; strain B.1.526, strain S:677H, and strain S:677P. In some embodiments, the one or more coronaviruses that cause the common cold are selected from: 229E alpha coronavirus, NL63 alpha coronavirus, OC43 beta coronavirus, and HKU1 beta coronavirus. In some embodiments, the conserved large sequences may be considered the 2 most highly conserved sequences of the identified large sequences in the alignment. In some embodiments, the conserved large sequences may be considered the 5 most highly conserved sequences of the identified large sequences in the alignment. In some embodiments, the conserved large sequences may be considered the 10 most highly conserved sequences of the identified large sequences in the alignment. In some embodiments, the conserved large sequences may be considered the 15 most highly conserved sequences of the identified large sequences in the alignment.

[0085] The present invention also features methods for preventing coronavirus disease. The method comprises administering to a subject a therapeutically effective amount of a pan-coronavirus recombinant vaccine composition according to the present invention, wherein the composition elicits an immune response in the subject and helps prevent coronavirus disease.

[0086] The present invention also features methods for preventing a coronavirus infection prophylactically in a subject. In some embodiments, the method comprises administering to the subject a

prophylactically effective amount of a pan-coronavirus recombinant vaccine composition according to the present invention, wherein the vaccine composition prevents coronavirus infection.

[0087] The present invention also features methods for eliciting an immune response in a subject, comprising administering to the subject a composition according to the present invention, wherein the vaccine composition elicits an immune response in the subject. The present invention also features methods comprising: administering to a subject a pan-coronavirus recombinant vaccine composition according to the present invention, wherein the composition prevents virus replication in the lungs, the brain, and other compartments where the virus replicates. The present invention also features methods comprising: administering to the subject a pan-coronavirus recombinant vaccine composition according to the present invention, wherein the composition prevents cytokine storm in the lungs, the brain, and other compartments where the virus replicates. The present invention also features methods comprising: administering to the subject a pan-coronavirus recombinant vaccine composition according to the present invention, wherein the composition prevents inflammation or inflammatory response in the lungs, the brain, and other compartments where the virus replicates. The present invention also features methods comprising: administering to the subject a pan-coronavirus recombinant vaccine composition according to the present invention, wherein the composition improves homing and retention of T cells in the lungs, the brain, and other compartments where the virus replicates. The present invention also features methods for preventing coronavirus disease in a subject; the method comprising: administering to the subject a pan-coronavirus recombinant vaccine composition according to the present invention, wherein the composition induces memory B and T cells. The present invention also features methods for prolonging an immune response induced by a pan-coronavirus recombinant vaccine and increasing T-cell migration to the lungs, the method comprising: co-expressing a T-cell attracting chemokine, a composition that promotes T cell proliferation, and a pan-coronavirus recombinant vaccine according to the present invention. The present invention also features methods for prolonging the retention of memory T-cell into the lungs induced by a pan coronavirus vaccine and increasing virus-specific tissue resident memory T-cells (TRM cells), the method comprising: co-expressing a T-cell attracting chemokine, a composition that promotes T cell proliferation, and a pan-coronavirus recombinant vaccine according to the present invention. The present invention also features methods comprising: administering to the subject a pan-coronavirus recombinant vaccine composition according to the present invention, wherein the composition prevents the development of mutation and variants of a coronavirus.

[0088] For the sake of brevity, it is noted that the vaccine compositions referred to in the aforementioned methods include the vaccine compositions previously discussed, the embodiments described below, and the embodiments in the figures.

[0089] In some embodiments, the vaccine composition is administered through an intravenous route (i.v.), an intranasal route (i.n.), or a sublingual route (s.l.) route.

[0090] In some embodiments, the vaccine composition is administered using an adenovirus or other appropriate delivery system.

[0091] As previously discussed, the composition herein may be used to prevent a coronavirus disease in a subject. The composition herein may be used to prevent a coronavirus infection prophylactically in a subject. The composition herein may be used to elicit an immune response in a subject. The term "subject" herein may refer to a human, a non-human primate, an animal such as a mouse, rat, cat, dog, other animal that is susceptible to coronavirus infection, or other animal used for preclinical modeling. The composition herein may prolong an immune response induced by the pan-coronavirus recombinant vaccine composition and increases T-cell migration to the lungs. In certain embodiments, the composition induces resident memory T cells (Trm). In some embodiments, the vaccine composition induces efficient and powerful protection against the coronavirus disease or infection. In some embodiments, the vaccine composition induces production of antibodies (Abs), CD4+ T helper (Th1) cells, and CD8+ cytotoxic T-cells (CTL). In some embodiments, the composition that promotes T cell proliferation helps to promote long term immunity. In some embodiments, the T-cell attracting chemokine helps pull T-cells from circulation into the lungs.

[0092] The present invention also features oligonucleotide compositions. For example, the present invention includes oligonucleotides disclosed in the sequence listings. The present invention also includes oligonucleotides in the form of antigen delivery systems. The present invention also includes oligonucleotides encoding the conserved large sequences disclosed herein. The present invention also includes oligonucleotide compositions comprising one or more oligonucleotides encoding any of the vaccine compositions according to the present invention. In some embodiments, the oligonucleotide comprises DNA. In some embodiments, the oligonucleotide comprises modified DNA. In some embodiments, the oligonucleotide comprises RNA. In some embodiments, the oligonucleotide comprises modified RNA. In some embodiments, the oligonucleotide comprises mRNA. In some embodiments, the oligonucleotide comprises modified mRNA.

[0093] The present invention also features peptide compositions. For example, the present invention includes peptides disclosed in the sequence listings. The present invention also includes peptide compositions comprising any of the vaccine compositions according to the present invention. The present invention also includes peptide compositions comprising any of the conserved large sequences according to the present invention.

[0094] For the sake of brevity, it is noted that the vaccine compositions referred to in the aforementioned oligonucleotide and peptide compositions include the vaccine compositions previously discussed, the embodiments described below, and the embodiments in the figures.

[0095] The present invention also features a pan-coronavirus recombinant vaccine composition comprising SEQ ID NO: 139 - 147 (Table 9).

[0096] The present invention also features a pan-coronavirus recombinant vaccine composition at least 99% identical to SEQ ID NO: 139 - 147 (Table 9).

[0097] The present invention also features a method comprising: administering a first pan-coronavirus

recombinant vaccine dose using a first delivery system, and administering a second vaccine dose using a second delivery system, wherein the first and second delivery system are different. In some embodiments, the first delivery system may comprise a RNA, a modified mRNA, or a peptide delivery system. In some embodiments, the second delivery system may comprise a RNA, a modified mRNA, or a peptide delivery system. In some embodiments, the peptide delivery system is an adenovirus. In some embodiments, the adenovirus delivery system is Ad26, Ad5, Ad35, or a combination thereof. In some embodiments, the peptide delivery system is a vesicular stomatitis virus (VSV) vector. In some embodiments, the second vaccine dose is administered 14 days after the first vaccine dose.

[0098] The present invention also features a method comprising: administering a pan-coronavirus recombinant vaccine composition according to the present invention; and administering at least one T-cell attracting chemokine after administering the pan-coronavirus recombinant vaccine composition. In some embodiments, the vaccine composition is administered via a RNA, a modified mRNA, or a peptide delivery system. In some embodiments, the T-cell attracting chemokine is administered via a RNA, a modified mRNA, or a peptide delivery system. In some embodiments, the peptide delivery system is an adenovirus. In some embodiments, the adenovirus delivery system is Ad26, Ad5, Ad35, or a combination thereof. In some embodiments, the peptide delivery system is a vesicular stomatitis virus (VSV) vector. In some embodiments, the T-cell attracting chemokine is administered 8 days after administering days after the vaccine composition. In some embodiments, the T-cell attracting chemokine is administered 14 days after administering days after the vaccine composition. In some embodiments, the T-cell attracting chemokine is administered 30 days after administering days after the vaccine composition. In some embodiments, the T-cell attracting chemokine is CCL5, CXCL9, CXCL10, CXCL11, or a combination thereof.

[0099] The present invention also features a method comprising: administering a pan-coronavirus recombinant vaccine composition according to the present invention; administering at least one T-cell attracting chemokine after administering the pan-coronavirus recombinant vaccine composition; and administering at least one cytokine after administering the T-cell attracting chemokine. In some embodiments, the vaccine composition is administered via a RNA, a modified mRNA, or a peptide delivery system. In some embodiments, the T-cell attracting chemokine is administered via a RNA, a modified mRNA, or a peptide delivery system. In some embodiments, the cytokine is administered via a RNA, a modified mRNA, or a peptide delivery system. In some embodiments, the peptide delivery system is an adenovirus. In some embodiments, the adenovirus delivery system is Ad26, Ad5, Ad35, or a combination thereof. In some embodiments, the peptide delivery system is a vesicular stomatitis virus (VSV) vector. In some embodiments, the T-cell attracting chemokine is administered 14 days after administering the vaccine composition. In some embodiments, the T-cell attracting chemokine is CCL5, CXCL9, CXCL10, CXCL11, or a combination thereof. In some embodiments, the cytokine is administered 10 days after administering the T-cell attracting chemokine. In some embodiments, the cytokine is IL-7, IL-15, IL2 or a combination thereof.

[00100] The present invention also features a method comprising: administering a pan-coronavirus

recombinant vaccine composition according to the present invention; administering one or more T-cell attracting chemokine after administering the pan-coronavirus recombinant vaccine composition; and administering one or more mucosal chemokine(s). In some embodiments, the vaccine composition is administered using an adenovirus. In some embodiments, the T-cell attracting chemokine is administered via a RNA, a modified mRNA, or a peptide delivery system, or other delivery system. In some embodiments, the mucosal chemokine is administered via a RNA, a modified mRNA, or a peptide delivery system, or other delivery system. In some embodiments, the adenovirus is Ad26, Ad5, Ad35, or a combination thereof. In some embodiments, the T-cell attracting chemokine is administered 14 days after administering the vaccine composition. In some embodiments, the T-cell attracting chemokine is CCL5, CXCL9, CXCL10, CXCL11, or a combination thereof. In some embodiments, the mucosal chemokine is administered 10 days after administering the T-cell attracting chemokine. In some embodiments, the mucosal chemokine is CCL25, CCL28, CXCL14, or CXCL17, or a combination thereof.

[00101] For the sake of brevity, it is noted that the vaccine compositions referred to in the aforementioned methods include the vaccine compositions previously discussed, the embodiments described below, and the embodiments in the figures.

[00102] As previously discussed, in some embodiments, the vaccine compositions are for use in humans. In some embodiments, the vaccine compositions are for use in animals, e.g. cats, dogs, etc. In some embodiments, the vaccine composition comprises human CXCL-11 and/or human IL-7 (or IL-15, IL-2). In some embodiments, the vaccine composition comprises animal CXCL-11 and/or animal IL-7 (or IL-15, IL-2).

[00103] The present invention includes vaccine compositions in the form of a rVSV-panCoV vaccine composition. The present invention includes vaccine compositions in the form of a rAdV-panCoV vaccine composition.

[00104] The present invention also includes nucleic acids for use in the vaccine compositions herein. The present invention also includes vectors for use in the vaccine compositions herein. The present invention also includes fusion proteins for use in the vaccine compositions herein. The present invention also includes immunogenic compositions for use in the vaccine compositions herein.

[00105] The vaccine compositions herein may be designed to elicit both high levels of virus-blocking and virus-neutralizing antibodies as well as CD4+ T cells and CD8+ T cells in adults 18 to 55 years. The vaccine compositions herein may be designed to elicit both high levels of virus-blocking and virus-neutralizing antibodies as well as CD4+ T cells and CD8+ T cells in adults 55 to 65 years of age. The vaccine compositions herein may be designed to elicit both high levels of virus-blocking and virus-neutralizing antibodies as well as CD4+ T cells and CD8+ T cells in adults 65 to 85 years of age. The vaccine compositions herein may be designed to elicit both high levels of virus-blocking and virus-neutralizing antibodies as well as CD4+ T cells and CD8+ T cells in adults 85 to 100 years of age. The vaccine compositions herein may be designed to elicit both high levels of virus-blocking and virus-neutralizing antibodies as well as CD4+ T cells and CD8+ T cells in children 12 to 18 years of age.

The vaccine compositions herein may be designed to elicit both high levels of virus-blocking and virus-neutralizing antibodies as well as CD4+ T cells and CD8+ T cells in children under 12 years of age.

[00106] The present invention is not limited to vaccine compositions. For example, in certain embodiments, one or more of the conserved large sequences are used for detecting coronavirus and/or diagnosing coronavirus infection.

[00107] As previously discussed, in some embodiments, the one or more conserved large sequences are highly conserved among human and animal coronaviruses. In some embodiments, the conserved large sequence is one that is among the most highly conserved large sequences identified in a sequence alignment and analysis of a particular number of coronavirus sequences. For example, the conserved large sequence may be the 2 most highly conserved large sequences identified. In some embodiments, the conserved large sequences may be the 5 most highly conserved large sequences identified. In some embodiments, the conserved large sequences may be the 8 most highly conserved large sequences identified. In some embodiments, the conserved large sequences may be the 10 most highly conserved large sequences identified. In some embodiments, the conserved large sequences may be the 15 most highly conserved large sequences identified. In some embodiments, the conserved large sequences may be the 20 most highly conserved large sequences identified. In some embodiments, the conserved large sequences may be the 30 most highly conserved large sequences identified. In some embodiments, the conserved large sequences may be the 40 most highly conserved large sequences identified. In some embodiments, the one or more conserved In some embodiments, the conserved large sequences may be the 5 most highly conserved large sequences identified are derived from at least one of SARS-CoV-2 protein. In some embodiments, the one or more conserved In some embodiments, the conserved large sequences may be the 5 most highly conserved large sequences identified are derived from one or more of: one or more SARS-CoV-2 human strains or variants in current circulation; one or more coronaviruses that has caused a previous human outbreak; one or more coronaviruses isolated from animals selected from a group consisting of bats, pangolins, civet cats, minks, camels, and other animal receptive to coronaviruses; or one or more coronaviruses that cause the common cold. In some embodiments, the one or more SARS-CoV-2 human strains or variants in current circulation are selected from: strain B.1.177; strain B.1.160, strain B.1.1.7; strain B.1.351; strain P.1; strain B.1.427/B.1.429; strain B.1.258; strain B.1.221; strain B.1.367; strain B.1.1.277; strain B.1.1.302; strain B.1.525; strain B.1.526, strain S.677H, and strain S.677P. In some embodiments, the one or more coronaviruses that cause the common cold are selected from: 229E alpha coronavirus, NL63 alpha coronavirus, OC43 beta coronavirus, and HKU1 beta coronavirus. In some embodiments, the vaccine composition is for humans. In some embodiments, the vaccine composition is for animals.

[00108] The present invention also features a method of producing a pan-coronavirus composition, the method comprising selecting at least one large sequence(s) according to the present invention and synthesizing one or more antigens comprising the selected large sequence(s). The present invention also features a method of producing a pan-coronavirus composition, the method comprising selecting at least one conserved large sequence(s); and synthesizing an antigen delivery system that encodes the selected

large sequence(s).

[00109] The present invention also includes a pan-coronavirus recombinant vaccine composition, the composition comprising one or more large sequences, each of the one or more large sequences comprises at least one of: whole spike protein or a portion thereof; one or more conserved coronavirus CD4+ T cell target epitope; and one or more conserved coronavirus CD8+ T cell target epitope; wherein at least one epitope is derived from a non-spike protein.

[00110] In some embodiments, the one or more conserved epitopes are highly conserved among human and animal coronaviruses. In some embodiments, the one or more conserved epitopes are derived from at least one of SARS-CoV-2 protein. In some embodiments, the composition comprises 2-20 CD8+ T cell target epitopes. In some embodiments, the composition comprises 2-20 CD4+ T cell target epitopes. In some embodiments, the one or more conserved coronavirus CD4+ T cell target epitopes selected from SEQ ID NO: 58-105 (ORF1a1350-1365, ORF1ab5019-5033, ORF612-26, ORF1ab6088-6102, ORF1ab6420-6434, ORF1a1801-1815, S1-13, E26-40, E20-34, M176-190, N388-403, ORF7a3-17, ORF7a1-15, ORF7b8-22, ORF7a98-112, and ORF81-15.). In some embodiments, the one or more conserved coronavirus CD8+ T cell target epitopes selected from SEQ ID NO: 106-138 (S287-317, S524-598, S601-640, S802-819, S888-909, S369-393, S440-501, S1133-1172, S329-363, and S13-37).

[00111] The present invention also features a pan-coronavirus recombinant vaccine composition, the composition comprising one or more large sequences, each of the one or more large sequences comprises at least one of: one or more conserved coronavirus B-cell target epitope; one or more conserved coronavirus CD4+ T cell target epitope; and/or one or more conserved coronavirus CD8+ T cell target epitope, wherein at least one epitope is derived from a non-spike protein.

[00112] In some embodiments, the one or more conserved epitopes are derived from at least one of SARS-CoV-2 proteins. In some embodiments, the composition comprises 2-20 CD8+ T cell target epitopes. In some embodiments, the composition comprises 2-20 CD4+ T cell target epitopes. In some embodiments, the one or more conserved coronavirus CD4+ T cell target epitopes selected from SEQ ID NO: 58-105 (ORF1a1350-1365, ORF1ab5019-5033, ORF612-26, ORF1ab6088-6102, ORF1ab6420-6434, ORF1a1801-1815, S1-13, E26-40, E20-34, M176-190, N388-403, ORF7a3-17, ORF7a1-15, ORF7b8-22, ORF7a98-112, and ORF81-15.).

[00113] In some embodiments, the one or more conserved coronavirus CD8+ T cell target epitopes selected from SEQ ID NO: 106-138 (S287-317, S524-598, S601-640, S802-819, S888-909, S369-393, S440-501, S1133-1172, S329-363, and S13-37).

[00114] In some embodiments, the one or more conserved coronavirus B-cell target epitopes selected from SEQ ID NO: 2-57 (S2-10, S1220-1228, S1000-1008, S958-966, E20-28, ORF1ab1675-1683, ORF1ab2363-2371, ORF1ab3013-3021, ORF1ab3183-3191, ORF1ab5470-5478, ORF1ab6749-6757, ORF7b26-34, ORF8a73-81, ORF103-11, and ORF105-13);

[00115] The present invention also features a pan-coronavirus recombinant vaccine composition, the

composition comprising an antigen delivery system encoding one or more large sequences, the large sequences comprise at least one of: one or more conserved coronavirus B-cell target epitopes; one or more conserved coronavirus CD4+ T cell target epitopes; and/or one or more conserved coronavirus CD8+ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.

[00116] In some embodiments, the antigen delivery system is an adenovirus-based antigen delivery system. In some embodiments, the adenovirus-based antigen delivery system is Ad26, Ad5, Ad35, or a combination thereof. In some embodiments, the antigen delivery system further encodes a T cell attracting chemokine. In some embodiments, the antigen delivery system further encodes a composition that promotes T cell proliferation. In some embodiments, the antigen delivery system further encodes a molecular adjuvant. In some embodiments, the large sequences are operatively linked to a lung-specific promoter.

[00117] In some embodiments, the one or more conserved coronavirus B-cell target epitopes selected from SEQ ID NO: 2-57 (S2-10, S1220-1228, S1000-1008, S958-966, E20-28, ORF1ab1675-1683, ORF1ab2363-2371, ORF1ab3013-3021, ORF1ab3183-3191, ORF1ab5470-5478, ORF1ab6749-6757, ORF7b26-34, ORF8a73-81, ORF103-11, and ORF105-13). In some embodiments, the one or more conserved coronavirus CD4+ T cell target epitopes selected from SEQ ID NO: 58-105 (ORF1a1350-1365, ORF1ab5019-5033, ORF612-26, ORF1ab6088-6102, ORF1ab6420-6434, ORF1a1801-1815, S1-13, E26-40, E20-34, M176-190, N388-403, ORF7a3-17, ORF7a1-15, ORF7b8-22, ORF7a98-112, and ORF81-15.). In some embodiments, the one or more conserved coronavirus CD8+ T cell target epitopes selected from SEQ ID NO: 106-138 (S287-317, S524-598, S601-640, S802-819, S888-909, S369-393, S440-501, S1133-1172, S329-363, and S13-37).

[00118] In some embodiments, the partial spike protein comprises a trimerized SARS-CoV-2 receptor-binding domain (RBD). In some embodiments, the whole spike protein or partial spike protein has an intact S1-S2 cleavage site. In some embodiments, the spike protein is stabilized with proline substitutions at amino acid positions 986 and 987.

[00119] The present invention also features a pan-coronavirus recombinant vaccine composition comprising one of SEQ ID NO: 139-147.

[00120] The present invention also includes the corresponding nucleic acid sequences for any of the protein sequences herein. The present invention also includes the corresponding protein sequences for any of the nucleic acid sequences herein.

[00121] Embodiments herein may comprise whole spike protein or a portion of spike protein. Whole spike protein and a portion thereof is not limited to a wild type or original sequence and may include spike protein or a portion thereof with one or more modifications and/or mutations, such as point mutations, deletions, etc., including the mutations described herein such as those for improving stability.

[00122] Embodiments of the present invention can be freely combined with each other if they are not mutually exclusive.

[00123] Any feature or combination of features described herein are included within the scope of the present invention provided that the features included in any such combination are not mutually inconsistent as will be apparent from the context, this specification, and the knowledge of one of ordinary skill in the art. Additional advantages and aspects of the present invention are apparent in the following detailed description and claims.

DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING(S)

[00124] The features and advantages of the present invention will become apparent from a consideration of the following detailed description presented in connection with the accompanying drawings in which:

[00125] FIG. 1 shows a schematic view of an example of a large sequence pan-coronavirus recombinant vaccine composition. Each large sequence in the recombinant vaccine composition may comprise epitopes. CD8+ T cell epitopes are shown with a square, CD4+ T cell epitopes are shown with a circle and B-cell epitopes are shown with a diamond. Each shape (square, circle, or diamond) may represent a variety of different epitopes and is not limited to a singular epitope. The multi-epitope pan-coronavirus vaccines are not limited to a specific combination of large sequences as shown. The large sequence pan-coronavirus vaccines may comprise a various number large sequences.

[00126] FIG. 2A shows an evolutionary comparison of genome sequences among beta-Coronavirus strains isolated from humans and animals. A phylogenetic analysis performed between SARS-CoV-2 strains (obtained from humans (*Homo Sapiens* (black)), along with the animal's SARS-like Coronaviruses genome sequence (SL-CoVs) sequences obtained from bats (*Rhinolophus affinis*, *Rhinolophus malayanus* (red)), pangolins (*Manis javanica* (blue)), civet cats (*Paguma larvata* (green)), and camels (*Camelus dromedaries* (Brown)). The included SARS-CoV/MERS-CoV strains are from previous outbreaks (obtained from humans (Urbani, MERS-CoV, OC43, NL63, 229E, HKU1-genotype-B), bats (WIV16, WIV1, YNLF-31C, Rs672, recombinant strains), camel (*Camelus dromedaries*, (KT368891.1, MN514967.1, KF917527.1, NC_028752.1), and civet (Civet007, A022, B039)). The human SARS-CoV-2 genome sequences are represented from six continents.

[00127] FIG. 2B shows shows an evolutionary analysis performed among the human-SARS-CoV-2 genome sequences reported from six continents and SARS-CoV-2 genome sequences obtained from bats (*Rhinolophus affinis*, *Rhinolophus malayanus*), and pangolins (*Manis javanica*)).

[00128] FIG. 3A shows lungs, heart, kidneys, intestines, brain, and testicles express ACE2 receptors and are targeted by SARS-CoV-2 virus. SARS-CoV-2 virus docks on the Angiotensin converting enzyme 2 (ACE2) receptor via spike surface protein.

[00129] FIG. 3B shows a System Biology Analysis approach utilized in the present invention.

[00130] FIG. 4 shows sequence homology analysis for SARS-CoV-2, common cold CoV strains, MERS, SARS-CoV-Urbani and animal CoVs with SARS-CoV-2 Wuhan Strain (Query strain; hCoV-19/batYN01). Five fragments SARS-CoV-2 genome were found to be highly conserved (1bp- 1580bp (fragment 1),

3547bp- 12830bp (fragment 2), 17472bp- 21156bp (fragment 3), 22584bp- 24682bp (fragment 4), and 26193bp- 27421bp (fragment 5).

[00131] FIG. 5 shows sequence homology analysis for fragment 1 (1bp- 1580bp) which comprises portions of ORF1a/b. The Query sequence (1-1580bp hCoV-19/batYN01) was BLAST against all the SARS-CoV-2 VOCs, human CoV strains, CoV strains from bats, pangolin, civet cat. 28 variants/strains were found with significant homology for this queried region.

[00132] FIG. 6 shows sequence homology analysis for fragment 2 (3547bp- 12830bp). The Query sequence (3547-12830 bp hCoV-19/batYN01) was BLAST against all the SARS-CoV-2 VOCs, human CoV strains, CoV strains from bats, pangolin, civet cats. 30 variants/strains were found with significant homology for this queried region.

[00133] FIG. 7 shows sequence homology analysis for fragment 3 (17472bp- 21156bp). The Query sequence (17472- 21156 bp hCoV-19/batYN01) was BLAST against all the SARS-CoV-2 VOCs, human CoV strains, CoV strains from bats, pangolin, civet cats. 29 variants/strains were found with significant homology for this queried region.

[00134] FIG. 8 shows sequence homology analysis for fragment 4 (22584bp- 24682bp) which comprises the spike protein. The Query sequence (22584- 24682 bp hCoV-19/batYN01) was BLAST against all the SARS-CoV-2 VOCs, human CoV strains, CoV strains from bats, pangolin, civet cats. 29 variants/strains were found with significant homology for this queried region.

[00135] FIG. 9 shows sequence homology analysis for fragment 5 (26193bp- 27421bp). The Query sequence (26193- 27421 bp hCoV-19/batYN01) was BLAST against all the SARS-CoV-2 VOCs, human CoV strains, CoV strains from bats, pangolin, civet cats. 31 variants/strains were found with significant homology for this queried region.

[00136] FIG. 10 shows a sequence homology analysis to screen conservancy of potential SARS-CoV-2-derived human CD8+ T cell epitopes. Shown are the comparison of sequence homology for the potential CD8+ T cell epitopes among 81,963 SARS-CoV-2 strains (that currently circulate in 190 countries on 6 continents), the 4 major "common cold" Coronaviruses that caused previous outbreaks (i.e. hCoV-OC43, hCoV-229E, hCoV-HKU1-Genotype B, and hCoV-NL63), and the SL-CoVs that were isolated from bats, civet cats, pangolins and camels. Epitope sequences highlighted in yellow present a high degree of homology among the currently circulating 81,963 SARS-CoV-2 strains and at least a 50% conservancy among two or more humans SARS-CoV strains from previous outbreaks, and the SL-CoV strains isolated from bats, civet cats, pangolins and camels, as described herein. Homo Sapiens- black, bats (*Rhinolophus affinis*, *Rhinolophus malayanus*-red), pangolins (*Manis javanica*-blue), civet cats (*Paguma larvata*-green), and camels (*Camelus dromedaries*-brown).

[00137] FIG. 11A shows docking of highly conserved SARS-CoV-2-derived human CD8+ T cell epitopes to HLA-A*02:01 molecules, e.g., docking of the 27 high-affinity CD8+ T cell binder peptides to the groove of HLA-A*02:01 molecules.

[00138] FIG. 11B shows a summary of the interaction similarity scores of the 27 high-affinity CD8+ T cell epitope peptides to HLA-A*02:01 molecules determined by protein-peptide molecular docking analysis. Black columns depict CD8+ T cell epitope peptides with high interaction similarity scores.

[00139] FIG. 12A shows an experimental design show CD8+ T cells are specific to highly conserved SARS-CoV-2 epitopes detected in COVID-19 patients and unexposed healthy individuals: PBMCs from HLA-A*02:01 positive COVID-19 patients (n = 30) and controls unexposed healthy individuals (n = 10) were isolated and stimulated overnight with 10 μ M of each of the 27 SARS-CoV-2-derived CD8+ T cell epitopes. The number of IFN- γ -producing cells were quantified using ELISpot assay.

[00140] FIG. 12B shows the results from FIG. 12A. Dotted lines represent threshold to evaluate the relative magnitude of the response: a mean SFCs between 25 and 50 correspond to a medium/intermediate response whereas a strong response is defined for a mean SFCs > 50.

[00141] FIG. 12C shows the results from experiments where PBMCs from HLA-A*02:01 positive COVID-19 patients were further stimulated for an additional 5 hours in the presence of mAbs specific to CD107a and CD107b, and Golgi-plug and Golgi-stop. Tetramers specific to Spike epitopes, CD107a/b and CD69 and TNF- expression were then measured by FACS. Representative FACS plot showing the frequencies of Tetramer+CD8+ T cells, CD107a/b+CD8+ T cells, CD69+CD8+ T cells and TNF-+CD8+ T cells following priming with a group of 4 Spike CD8+ T cell epitope peptides. Average frequencies of tetramer+CD8+ T cells, CD107a/b+CD8+ T cells, CD69+CD8+ T cells and TNF-+CD8+ T cells.

[00142] FIG. 13A shows a timeline of immunization and immunological analyses for experiments testing the immunogenicity of genome-wide identified human SARS-CoV-2 CD8+ T epitopes in HLA-A*02:01/HLA-DRB1 double transgenic mice. Eight groups of age-matched HLA-A*02:01 transgenic mice (n = 3) were immunized subcutaneously, on days 0 and 14, with a mixture of four SARS-CoV-2-derived human CD8+ T cell peptide epitopes mixed with PADRE CD4+ T helper epitope, delivered in alum and CpG1826 adjuvants. As a negative control, mice received adjuvants alone (mock-immunized).

[00143] FIG. 13B shows the gating strategy used to characterize spleen-derived CD8+ T cells. Lymphocytes were identified by a low forward scatter (FSC) and low side scatter (SSC) gate. Singlets were selected by plotting forward scatter area (FSC-A) vs. forward scatter height (FSC-H). CD8 positive cells were then gated by the expression of CD8 and CD3 markers.

[00144] FIG. 13C shows a representative ELISpot images (left panel) and average frequencies (right panel) of IFN- γ -producing cell spots from splenocytes (106 cells/well) stimulated for 48 hours with 10 μ M of 10 immunodominant CD8+ T cell peptides and 1 subdominant CD8+ T cell peptide out of the total pool of 27 CD8+ T cell peptides derived from SARS-CoV-2 structural and non-structural proteins. The number on the top of each ELISpot image represents the number of IFN- γ -producing spot forming T cells (SFC) per one million splenocytes.

[00145] FIG. 13D shows a representative FACS plot (left panel) and average frequencies (right panel) of IFN- γ and TNF- production by, and CD107a/b and CD69 expression on 10 immunodominant CD8+ T cell peptides and 1 subdominant CD8+ T cell peptide out of the total pool of 27 CD8+ T cell peptides derived from SARS-CoV-2 structural and non-structural proteins determined by FACS. Numbers indicate frequencies of IFN- γ +CD8+ T cells, CD107+CD8+ T cells, CD69+CD8+ T cells and TNF-+CD8+ T cells, detected in 3 immunized mice.

[00146] FIG. 14 shows the SARS-CoV/SARS-CoV-2 genome encodes two large non-structural genes ORF1a (green) and ORF1b (gray), encoding 16 non-structural proteins (NSP1– NSP16). The genome encodes at least six accessory proteins (shades of light grey) that are unique to SARS-CoV/SARS-CoV-2 in terms of number, genomic organization, sequence, and function. The common SARS-CoV, SARS-CoV-2 and SL-CoVs-derived human B (blue), CD4+ (green) and CD8+ (black) T cell epitopes are shown. Structural and non-structural open reading frames utilized in this study were from SARS-CoV-2-Wuhan-Hu-1 strain (NCBI accession number MN908947.3, SEQ ID NO: 1). The amino acid sequence of the SARS-CoV-2-Wuhan-Hu-1 structural and non-structural proteins was screened for human B, CD4+ and CD8+ T cell epitopes using different computational algorithms as described herein. Shown are genome-wide identified SARS-CoV-2 human B cell epitopes (in blue), CD4+ T cell epitopes (in green), CD8+ T cell epitopes (in black) that are highly conserved between human and animal Coronaviruses.

[00147] FIG. 15 shows the identification of highly conserved potential SARS-CoV-2-derived human CD4+ T cell epitopes that bind with high affinity to HLA-DR molecules: Out of a total of 9,594 potential HLA-DR-restricted CD4+ T cell epitopes from the whole genome sequence of SARS-CoV-2-Wuhan-Hu-1 strain (MN908947.3), 16 epitopes that bind with high affinity to HLA-DRB1 molecules were selected. The conservancy of the 16 CD4+ T cell epitopes was analyzed among human and animal Coronaviruses. Shown are the comparison of sequence homology for the 16 CD4+ T cell epitopes among 81,963 SARS-CoV-2 strains (that currently circulate in 6 continents), the 4 major “common cold” Coronaviruses that caused previous outbreaks (i.e. hCoV-OC43, hCoV-229E, hCoV-HKU1, and hCoV-NL63), and the SL-CoVs that were isolated from bats, civet cats, pangolins and camels. Epitope sequences highlighted in green present high degree of homology among the currently circulating 81,963 SARS-CoV-2 strains and at least a 50% conservancy among two or more humans SARS-CoV strains from previous outbreaks, and the SL-CoV strains isolated from bats, civet cats, pangolins and camels, as described in Materials and Methods. Homo Sapiens- black, bats (*Rhinolophus affinis*, *Rhinolophus malayanus* -red), pangolins (*Manis javanica*-blue), civet cats (*Paguma larvata*-green), and camels (*Camelus dromedaries*-brown).

[00148] FIG. 16A the molecular docking of highly conserved SARS-CoV-2 CD4+ T cell epitopes to HLA-DRB1 molecules. Molecular docking of 16 CD4+ T cell epitopes, conserved among human SARS-CoV-2 strains, previous humans SARS/MERS-CoV and bat SL-CoVs into the groove of the HLA-DRB1 protein crystal structure (PDB accession no: 4UQ3) was determined using the GalaxyPepDock server. The 16 CD4+ T cell epitopes are promiscuous restricted to HLA-DRB1*01:01, HLA-DRB1*11:01, HLA-DRB1*15:01, HLA-DRB1*03:01 and HLA-DRB1*04:01 alleles. The CD4+ T cell

peptides are shown in ball and stick structures, and the HLA-DRB1 protein crystal structure is shown as a template. The prediction accuracy is estimated from a linear model as the relationship between the fraction of correctly predicted binding site residues and the template-target similarity measured by the protein structure similarity score (TM score) and interaction similarity score (SInter) obtained by linear regression. SInter shows the similarity of the amino acids of the CD8+ T cell peptides aligned to the contacting residues in the amino acids of the HLA-DRB1 template structure.

[00149] FIG. 16B shows histograms representing interaction similarity score of CD4+ T cells specific epitopes observed from the protein-peptide molecular docking analysis.

[00150] FIG. 17A shows an experimental design to show CD4+ T cells are specific to highly conserved SARS-CoV-2 epitopes detected in COVID-19 patients and unexposed healthy individuals: PBMCs from HLA-DRB1 positive COVID-19 patients (n = 30) and controls unexposed healthy individuals (n = 10) were isolated and stimulated for 48 hrs. with 10 μ M of each of the 16 SARS-CoV-2-derived CD4+ T cell epitopes. The number of IFN--producing cells were quantified using ELISpot assay.

[00151] FIG. 17B shows the results from FIG. 17A. Dotted lines represent a threshold to evaluate the relative magnitude of the response: a mean SFCs between 25 and 50 correspond to a medium/intermediate response, whereas a strong response is defined for a mean SFCs > 50. PBMCs from HLA-DRB1-positive COVID-19 patients

[00152] FIG. 17C shows the results from further stimulating for an additional 5 hours in the presence of mAbs specific to CD107a and CD107b, and Golgi-plug and Golgi-stop. Tetramers specific to two Spike epitopes, CD107a/b and CD69 and TNF-alpha expression were then measured by FACS. Representative FACS plot showing the frequencies of Tetramer+CD4+ T cells, CD107a/b+CD4+ T cells, CD69+CD4+ T cells and TNF-+CD4+ T cells following priming with a group of 2 Spike CD4+ T cell epitope peptides. Average frequencies are shown for tetramer+CD4+ T cells, CD107a/b+CD4+ T cells, CD69+CD4+ T cells and TNF-+CD4+ T cells.

[00153] FIG. 18A shows a timeline of immunization and immunological analyses for testing immunogenicity of genome-wide identified human SARS-CoV-2 CD4+ T epitopes in HLA-A*02:01/HLA-DRB1 double transgenic mice. Four groups of age-matched HLA-DRB1 transgenic mice (n = 3) were immunized subcutaneously, on days 0 and 14, with a mixture of four SARS-CoV-2-derived human CD4+ T cell peptide epitopes delivered in alum and CpG1826 adjuvants. As a negative control, mice received adjuvants alone (mock-immunized).

[00154] FIG. 18B shows the gating strategy used to characterize spleen-derived CD4+ T cells. CD4 positive cells were gated by the CD4 and CD3 expression markers.

[00155] FIG. 18C shows the representative ELISpot images (left panel) and average frequencies (right panel) of IFN- γ -producing cell spots from splenocytes (106 cells/well) stimulated for 48 hours with 10 μ M of 7 immunodominant CD4+ T cell peptides and 1 subdominant CD4+ T cell peptide out of the total pool of 16 CD4+ T cell peptides derived from SARS-CoV-2 structural and non-structural proteins. The number

of IFN- γ -producing spot forming T cells (SFC) per one million of total cells is presented on the top of each ELISpot image.

[00156] FIG. 18D shows the representative FACS plot (left panel) and average frequencies (right panel) show IFN- γ and TNF- α -production by, and CD107a/b and CD69 expression on 7 immunodominant CD4+ T cell peptides and 1 subdominant CD4+ T cell peptide out of the total pool of 16 CD4+ T cell peptides derived from SARS-CoV-2 determined by FACS. The numbers indicate percentages of IFN- γ +CD4+ T cells, CD107+CD4+ T cells, CD69+CD4+ T cells and TNF- α +CD4+ T cells detected in 3 immunized mice.

[00157] FIG. 19 shows the conservation of Spike-derived B cell epitopes among human, bat, civet cat, pangolin, and camel coronavirus strains: Multiple sequence alignment performed using ClustalW among 29 strains of SARS coronavirus (SARS-CoV) obtained from human, bat, civet, pangolin, and camel. This includes 7 human SARS/MERS-CoV strains (SARS-CoV-2-Wuhan (MN908947.3), SARS-HCoV-Urbani (AY278741.1), CoV-HKU1-Genotype-B (AY884001), CoV-OC43 (KF923903), CoV-NL63 (NC005831), CoV-229E (KY983587), MERS (NC019843)); 8 bat SARS-CoV strains (BAT-SL-CoV-WIV16 (KT444582), BAT-SL-CoV-WIV1 (KF367457.1), BAT-SL-CoV-YNLF31C (KP886808.1), BAT-SARS-CoV-RS672 (FJ588686.1), BAT-CoV-RATG13 (MN996532.1), BAT-CoV-YN01 (EPIISL412976), BAT-CoV-YN02 (EPIISL412977), BAT-CoV-19-ZXC21 (MG772934.1)); 3 Civet SARS-CoV strains (SARS-CoV-Civet007 (AY572034.1), SARS-CoV-A022 (AY686863.1), SARS-CoV-B039 (AY686864.1)); 9 pangolin SARS-CoV strains (PCoV-GX-P2V(MT072864.1), PCoV-GX-P5E(MT040336.1), PCoV-GX-P5L (MT040335.1), PCoV-GX-P1E (MT040334.1), PCoV-GX-P4L (MT040333.1), PCoV-MP789 (MT084071.1), PCoV-GX-P3B (MT072865.1), PCoV-Guangdong-P2S (EPIISL410544), PCoV-Guangdong (EPIISL410721)); 4 camel SARS-CoV strains (Camel-CoV-HKU23 (KT368891.1), DcCoV-HKU23 (MN514967.1), MERS-CoV-Jeddah (KF917527.1), Riyadh/RV141 (NC028752.1)) and 1 recombinant strain (FJ211859.1)). Regions highlighted with blue color represent the sequence homology. The B cell epitopes, which showed at least 50% conservancy among two or more strains of the SARS Coronavirus or possess receptor-binding domain (RBD) specific amino acids were selected as candidate epitopes.

[00158] FIG. 20A shows the docking of SARS-CoV-2 Spike glycoprotein-derived B cell epitopes to human ACE2 receptor, e.g., molecular docking of 22 B-cell epitopes, identified from the SARS-CoV-2 Spike glycoprotein, with ACE2 receptors. B cell epitope peptides are shown in ball and stick structures whereas the ACE2 receptor protein is shown as a template. S471-501 and S369-393 peptide epitopes possess receptor binding domain region specific amino acid residues. The prediction accuracy is estimated from a linear model as the relationship between the fraction of correctly predicted binding site residues and the template-target similarity measured by the protein structure similarity score and interaction similarity score (SI_{inter}) obtained by linear regression. SI_{inter} shows the similarity of amino acids of the B-cell peptides aligned to the contacting residues in the amino acids of the ACE2 template structure. Higher SI_{inter} score represents a more significant binding affinity among the ACE2 molecule and B-cell peptides.

[00159] FIG. 20B shows the summary of the interaction similarity score of 22 B cells specific epitopes observed from the protein-peptide molecular docking analysis. B cell epitopes with high interaction similarity scores are indicated in black.

[00160] FIG. 21A shows the timeline of immunization and immunological analyses for testing to show IgG antibodies are specific to SARS-CoV-2 Spike protein-derived B-cell epitopes in immunized B6 mice and in convalescent COVID-19 patients. A total of 22 SARS-CoV-2 derived B-cell epitope peptides selected from SARS-CoV-2 Spike protein and tested in B6 mice were able to induce antibody responses. Four groups of age-matched B6 mice ($n = 3$) were immunized subcutaneously, on days 0 and 14, with a mixture of 4 or 5 SARS-CoV-2 derived B-cell peptide epitopes emulsified in alum and CpG1826 adjuvants. Alum/CpG1826 adjuvants alone were used as negative controls (mock-immunized).

[00161] FIG. 21B shows the frequencies of IgG-producing CD3(-)CD138(+)B220(+) plasma B cells were determined in the spleen of immunized mice by flow cytometry. For example, FIG. 21B shows the gating strategy was as follows: Lymphocytes were identified by a low forward scatter (FSC) and low side scatter (SSC) gate. Singlets were selected by plotting forward scatter area (FSC-A) versus forward scatter height (FSC-H). B cells were then gated by the expression of CD3(-) and B220(+) cells and CD138 expression on plasma B cells determined.

[00162] FIG. 21C shows the frequencies of IgG-producing CD3(-)CD138(+)B220(+) plasma B cells were determined in the spleen of immunized mice by flow cytometry. For example, FIG. 21C shows a representative FACS plot (left panels) and average frequencies (right panel) of plasma B cells detected in spleen of immunized mice. The percentages of plasma CD138(-)B220(+)B cells are indicated on the top left of each dot plot.

[00163] FIG. 21D shows SARS-CoV-2 derived B-cell epitopes-specific IgG responses were quantified in immune serum, 14 days post-second immunization (i.e. day 28), by ELISpot (Number of IgG(+)Spots). Representative ELISpot images (left panels) and average frequencies (right panel) of anti-peptide specific IgG-producing B cell spots (1×10^6 splenocytes/well) following 4 days in vitro B cell polyclonal stimulation with mouse Poly-S (Immunospot). The top/left of each ELISpot image shows the number of IgG-producing B cells per half a million cells. ELISA plates were coated with each individual immunizing peptide.

[00164] FIG. 21E shows the B-cell epitopes-specific IgG concentrations ($\mu\text{g/mL}$) measured by ELISA in levels of IgG detected in peptide-immunized B6 mice, after subtraction of the background measured from mock-vaccinated mice. The dashed horizontal line indicates the limit of detection.

[00165] FIG. 21F and FIG. 21G show the B-cell epitopes-specific IgG concentrations ($\mu\text{g/mL}$) measured by ELISA in Level of IgG specific to each of the 22 Spike peptides detected SARS-CoV-2 infected patients ($n=40$), after subtraction of the background measured from healthy non-exposed individuals ($n=10$). Black bars and gray bars show high and medium immunogenic B cell peptides, respectively. The dashed horizontal line indicates the limit of detection.

[00166] FIG. 22 shows an example of a whole spike protein comprising mutations including 6 proline mutations. The 6 proline mutations comprise single point mutations F817P, A892P, A899P, A942P, K986P and V987P. Additionally, the spike protein comprises a 682-QQAQ-685 mutation of the furin cleavage site for protease resistance. In some embodiments, the K986P and V987P Mutations allow for perfusion stabilization. FIG. 22 also shows the following sequences: MFVFLVLLPLVSS (SEQ ID NO: 188), ATGTTCTGTTCTCCTGGTGCTGCTGCCCTGGTGAGCAGC (SEQ ID NO: 175), CAGCAGGCCAG (SEQ ID NO: 189), and CCCCCC (SEQ ID NO:190).

[00167] FIG. 23 shows non-limiting examples of how the large sequences of the compositions described herein may be arranged.

[00168] FIG. 24 shows a schematic representation of a prototype Coronavirus vaccine of the present invention. The present invention is not limited to the prototype coronavirus vaccines as shown.

[00169] FIG. 25A shows a non-limiting example of a method for delivering the vaccine composition described herein using a “prime/pull” regimen in humans. The method comprises administering a pan-coronavirus recombinant vaccine composition and further administering at least one T-cell attracting chemokine (e.g. CXCL11) after administering the pan-coronavirus recombinant vaccine composition.

[00170] FIG. 25B shows a non-limiting example of a method for delivering the vaccine composition described herein using a “prime/boost” regimen in humans. The method comprises administering a first composition, e.g., a first pan-coronavirus recombinant vaccine composition dose using a first delivery system and further administering a second composition, e.g., a second vaccine composition dose using a second delivery system. In some embodiments, the first delivery system and the second delivery system are different.

[00171] FIG. 25C shows a non-limiting example of a method for delivering the vaccine composition described herein using a “prime/pull/keep” regimen in humans to increase the size and maintenance of lung-resident B-cells, CD4+ T cells and CD8+ T cells to protect against SARS-CoV-2. The method comprises administering a pan-coronavirus recombinant vaccine composition and administering at least one T-cell attracting chemokine (e.g. CXCL11 or CXCL17) after administering the pan-coronavirus recombinant vaccine composition.

[00172] FIG. 25D shows a non-limiting example of a method for delivering the vaccine composition described herein using a “prime/pull/boost” regimen in humans to increase the size and maintenance of lung-resident B-cells, CD4+ T cells and CD8+ T cells to protect against SARS-CoV-2. The method comprises administering a pan-coronavirus recombinant vaccine composition and administering at least one T-cell attracting chemokine (e.g. CXCL11 or CXCL17) after administering the pan-coronavirus recombinant vaccine composition. The method further comprises administering at least one cytokine after administering the T-cell attracting chemokine (e.g. IL-7, IL-5, or IL-2).

[00173] FIG. 26A shows a non-limiting example of a method for delivering the vaccine composition described herein using a “prime/pull” regimen in domestic animals (e.g. cats or dogs). The method

comprises administering a pan-coronavirus recombinant vaccine composition and further administering at least one T-cell attracting chemokine (e.g. CXCL11) after administering the pan-coronavirus recombinant vaccine composition.

[00174] FIG. 26B shows a non-limiting example of a method for delivering the vaccine composition described herein using a "prime/boost" regimen in domestic animals (e.g. cats or dogs). The method comprises administering a first composition, e.g. a first pan-coronavirus recombinant vaccine composition dose using a first delivery system and further administering a second composition, e.g., a second vaccine composition dose using a second delivery system. In some embodiments, the first delivery system and the second delivery system are different.

[00175] FIG. 26C shows a non-limiting example of a method for delivering the vaccine composition described herein using a "prime/pull/keep" regimen in domestic animals (e.g. cats or dogs) to increase the size and maintenance of lung-resident B-cells, CD4+ T cells and CD8+ T cells to protect against SARS-CoV-2. The method comprises administering a pan-coronavirus recombinant vaccine composition and administering at least one T-cell attracting chemokine (e.g. CXCL11 or CXCL17) after administering the pan-coronavirus recombinant vaccine composition.

[00176] FIG. 26D shows a non-limiting example of a method for delivering the vaccine composition described herein using a "prime/pull/boost" regimen in domestic animals (e.g. cats or dogs) to increase the size and maintenance of lung-resident B-cells, CD4+ T cells and CD8+ T cells to protect against SARS-CoV-2. The method comprises administering a pan-coronavirus recombinant vaccine composition and administering at least one T-cell attracting chemokine (e.g. CXCL11 or CXCL17) after administering the pan-coronavirus recombinant vaccine composition. The method further comprises administering at least one cytokine after administering the T-cell attracting chemokine (e.g. IL-7, IL-5, or IL-2).

TERMS

[00177] Unless otherwise explained, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which a disclosed invention belongs. The singular terms "a," "an," and "the" include plural referents unless context clearly indicates otherwise. Similarly, the word "or" is intended to include "and" unless the context clearly indicates otherwise. The term "comprising" means that other elements can also be present in addition to the defined elements presented. The use of "comprising" indicates inclusion rather than limitation. Stated another way, the term "comprising" means "including principally, but not necessary solely". Furthermore, variation of the word "comprising", such as "comprise" and "comprises", have correspondingly the same meanings. In one respect, the technology described herein related to the herein described compositions, methods, and respective component(s) thereof, as essential to the invention, yet open to the inclusion of unspecified elements, essential or not ("comprising").

[00178] Suitable methods and materials for the practice and/or testing of embodiments of the disclosure are described below. Such methods and materials are illustrative only and are not intended to be limiting.

Other methods and materials similar or equivalent to those described herein can be used. For example, conventional methods well known in the art to which the disclosure pertains are described in various general and more specific references, including, for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, 1989; Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 3d ed., Cold Spring Harbor Press, 2001; Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing Associates, 1992 (and Supplements to 2000); Ausubel *et al.*, *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, 4th ed., Wiley & Sons, 1999; Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1990; and Harlow and Lane, *Using Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1999, *Gene Expression Technology* (Methods in Enzymology, Vol. 185, edited by D. Goeddel, 1991. Academic Press, San Diego, Calif.), "Guide to Protein Purification" in *Methods in Enzymology* (M. P. Deutscher, ed., (1990) Academic Press, Inc.); *PCR Protocols: A Guide to Methods and Applications* (Innis, *et al.* 1990. Academic Press, San Diego, Calif.), *Culture of Animal Cells: A Manual of Basic Technique*, 2nd Ed. (R. I. Freshney. 1987. Liss, Inc. New York, N.Y.), *Gene Transfer and Expression Protocols*, pp. 109-128, ed. E. J. Murray, The Humana Press Inc., Clifton, N.J.), and the Ambion 1998 Catalog (Ambion, Austin, Tex.), the disclosures of which are incorporated in their entirety herein by reference.

[00179] Although methods and materials similar or equivalent to those described herein can be used to practice or test the disclosed technology, suitable methods and materials are described below. The materials, methods, and examples are illustrative only and not intended to be limiting.

[00180] As used herein, the terms "immunogenic protein, polypeptide, or peptide" or "antigen" refer to polypeptides or other molecules (or combinations of polypeptides and other molecules) that are immunologically active in the sense that once administered to the host, it is able to evoke an immune response of the humoral and/or cellular type directed against the protein. In embodiments, the protein fragment has substantially the same immunological activity as the total protein. Thus, a protein fragment according to the disclosure can comprises or consists essentially of or consists of at least one epitope or antigenic determinant. An "immunogenic" protein or polypeptide, as used herein, may include the full-length sequence of the protein, analogs thereof, or immunogenic fragments thereof. "Immunogenic fragment" refers to a fragment of a protein which includes one or more epitopes and thus elicits the immunological response described above.

[00181] Synthetic antigens are also included within the definition, for example, poly-epitopes, flanking epitopes, and other recombinant or synthetically derived antigens. Immunogenic fragments for purposes of the disclosure may feature at least about 1 amino acid, at least about 3 amino acids, at least about 5 amino acids, at least about 10-15 amino acids, or about 15-25 amino acids or more amino acids, of the molecule. There is no critical upper limit to the length of the fragment, which could comprise nearly the full-length of the protein sequence, or the full-length of the protein sequence, or even a fusion protein comprising at least one epitope of the protein.

[00182] As used herein, the term "epitope" refers to the site on an antigen or hapten to which specific B cells and/or T cells respond. The term is also used interchangeably with "antigenic determinant" or "antigenic determinant site". Antibodies that recognize the same epitope can be identified in a simple immunoassay showing the ability of one antibody to block the binding of another antibody to a target antigen.

[00183] As used herein, the term "immunological response" to a composition or vaccine refers to the development in the host of a cellular and/or antibody-mediated immune response to a composition or vaccine of interest. Usually, an "immunological response" includes but is not limited to one or more of the following effects: the production of antibodies, B cells, helper T cells, and/or cytotoxic T cells, directed specifically to an antigen or antigens included in the composition or vaccine of interest. The host may display either a therapeutic or protective immunological response so resistance to new infection will be enhanced and/or the clinical severity of the disease reduced. Such protection will be demonstrated by either a reduction or lack of symptoms normally displayed by an infected host, a quicker recovery time and/or a lowered viral titer in the infected host.

[00184] As used herein, the term "variant" refers to a substantially similar sequence. For polynucleotides, a variant comprises a deletion and/or addition and/or change of one or more nucleotides at one or more sites within the native polynucleotide and/or a substitution of one or more nucleotides at one or more sites in the native polynucleotide. As used herein, a "native" polynucleotide or polypeptide comprises a naturally occurring nucleotide sequence or an amino acid sequence, respectively. Variants of a particular polynucleotide of the disclosure (e.g., the reference polynucleotide) can also be evaluated by comparison of the percent sequence identity between the polypeptide encoded by a variant polynucleotide and the polypeptide encoded by the reference polynucleotide. "Variant" protein is intended to mean a protein derived from the native protein by deletion or addition of one or more amino acids at one or more sites in the native protein and/or substitution of one or more amino acids at one or more sites in the native protein. Variant proteins encompassed by the present disclosure are biologically active, that is they have the ability to elicit an immune response.

[00185] The HLA-DR/HLA-A*0201/hACE2 triple transgenic mouse model referred to herein is a novel susceptible animal model for pre-clinical testing of human COVID-19 vaccine candidates derived from crossing ACE2 transgenic mice with the unique HLA-DR/HLA-A*0201 double transgenic mice. ACE2 transgenic mice are a hACE2 transgenic mouse model expressing human ACE2 receptors in the lung, heart, kidney and intestine (Jackson Laboratory, Bar Harbor, ME). The HLA-DR/HLA-A*0201 double transgenic mice are "humanized" HLA double transgenic mice expressing Human Leukocyte Antigen HLA-A*0201 class I and HLA DR*0101 class II in place of the corresponding mouse MHC molecules (which are knocked out). The HLA-A*0201 haplotype was chosen because it is highly represented (> 50%) in the human population, regardless of race or ethnicity. The HLA-DR/HLA-A*0201/hACE2 triple transgenic mouse model is a "humanized" transgenic mouse model and has three advantages: (1) it is susceptible to human SARS-CoV2 infection; (2) it develops symptoms similar to those seen in COVID-19

in humans; and (3) it develops CD4⁺ T cells and CD8⁺ T cells response to human epitopes. The novel HLA-DR/HLA-A*0201/hACE2 triple transgenic mouse model of the present invention may be used in the pre-clinical testing of safety, immunogenicity and protective efficacy of the human multi-epitope COVID-19 vaccine candidates of the present invention.

[00186] As used herein, the terms "treat" or "treatment" or "treating" refers to both therapeutic treatment and prophylactic or preventative measures, wherein the object is to prevent or slow the development of the disease, such as slow down the development of a disorder, or reducing at least one adverse effect or symptom of a condition, disease or disorder, e.g., any disorder characterized by insufficient or undesired organ or tissue function. Treatment is generally "effective" if one or more symptoms or clinical markers are reduced as that term is defined herein. Alternatively, a treatment is "effective" if the progression of a disease is reduced or halted. That is, "treatment" includes not just the improvement of symptoms or decrease of markers of the disease, but also a cessation or slowing of progress or worsening of a symptom that would be expected in absence of treatment. Beneficial or desired clinical results include, but are not limited to, alleviation of one or more symptom(s), diminishment of extent of disease, stabilized (e.g., not worsening) state of disease, delay or slowing of disease progression, amelioration or palliation of the disease state, and remission (whether partial or total), whether detectable or undetectable. "Treatment" can also mean prolonging survival as compared to expected survival if not receiving treatment. "Treatment" also includes ameliorating a disease, lessening the severity of its complications, preventing it from manifesting, preventing it from recurring, merely preventing it from worsening, mitigating an inflammatory response included therein, or a therapeutic effort to affect any of the aforementioned, even if such therapeutic effort is ultimately unsuccessful.

[00187] As used herein, the term "carrier" or "pharmaceutically acceptable carrier" or "pharmaceutically acceptable vehicle" refers to any appropriate or useful carrier or vehicle for introducing a composition to a subject. Pharmaceutically acceptable carriers or vehicles may be conventional but are not limited to conventional vehicles. For example, E. W. Martin, *Remington's Pharmaceutical Sciences*, Mack Publishing Co., Easton, PA, 15th Edition (1975) and D. B. Troy, ed. *Remington: The Science and Practice of Pharmacy*, Lippincott Williams & Wilkins, Baltimore MD and Philadelphia, PA, 21st Edition (2006) describe compositions and formulations suitable for pharmaceutical delivery of one or more therapeutic compounds or molecules. Carriers (e.g., pharmaceutical carriers, pharmaceutical vehicles, pharmaceutical compositions, pharmaceutical molecules, etc.) are materials generally known to deliver molecules, proteins, cells and/or drugs and/or other appropriate material into the body. In general, the nature of the carrier will depend on the nature of the composition being delivered as well as the particular mode of administration being employed. In addition to biologically-neutral carriers, pharmaceutical compositions administered may contain minor amounts of non-toxic auxiliary substances, such as wetting or emulsifying agents, preservatives, and pH buffering agents and the like. Patents that describe pharmaceutical carriers include, but are not limited to: U.S. Patent No. 6,667,371; U.S. Patent No. 6,613,355; U.S. Patent No. 6,596,296; U.S. Patent No. 6,413,536; U.S. Patent No. 5,968,543; U.S. Patent No. 4,079, 038; U.S. Patent No. 4,093,709; U.S. Patent No. 4,131,648; U.S. Patent No. 4,138,344; U.S.

Patent No. 4,180,646; U.S. Patent No. 4,304,767; U.S. Patent No. 4,946,931, the disclosures of which are incorporated in their entirety by reference herein. The carrier may, for example, be solid, liquid (e.g., a solution), foam, a gel, the like, or a combination thereof. In some embodiments, the carrier comprises a biological matrix (e.g., biological fibers, etc.). In some embodiments, the carrier comprises a synthetic matrix (e.g., synthetic fibers, etc.). In certain embodiments, a portion of the carrier may comprise a biological matrix and a portion may comprise synthetic matrix.

[00188] As used herein "coronavirus" may refer to a group of related viruses such as but not limited to severe acute respiratory syndrome (SARS), middle east respiratory syndrome (MERS), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). All the coronaviruses cause respiratory tract infection that range from mild to lethal in mammals. Several non-limiting examples of Coronavirus strains are described herein.

[00189] As used herein, "severe acute respiratory syndrome coronavirus 2 (SARS-CoV2)" is a betacoronavirus that causes Coronavirus Disease 19 (COVID-19).

[00190] A "subject" is an individual and includes, but is not limited to, a mammal (e.g., a human, horse, pig, rabbit, dog, sheep, goat, non-human primate, cow, cat, guinea pig, or rodent), a fish, a bird, a reptile or an amphibian. The term does not denote a particular age or sex. Thus, adult and newborn subjects, as well as fetuses, whether male or female, are intended to be included. A "patient" is a subject afflicted with a disease or disorder. The term "patient" includes human and veterinary subjects

[00191] The terms "administering", and "administration" refer to methods of providing a pharmaceutical preparation to a subject. Such methods are well known to those skilled in the art and include, but are not limited to, administering the compositions orally, parenterally (e.g., intravenously and subcutaneously), by intramuscular injection, by intraperitoneal injection, intrathecally, transdermally, extracorporeally, topically or the like.

[00192] A composition can also be administered by topical intranasal administration (intranasally) or administration by inhalant. As used herein, "topical intranasal administration" means delivery of the compositions into the nose and nasal passages through one or both of the nares and can comprise delivery by a spraying mechanism (device) or droplet mechanism (device), or through aerosolization of the composition. Administration of the compositions by inhalant can be through the nose or mouth via delivery by a spraying or droplet mechanism. As used herein, "an inhaler" can be a spraying device or a droplet device for delivering a composition comprising the vaccine composition, in a pharmaceutically acceptable carrier, to the nasal passages and the upper and/or lower respiratory tracts of a subject. Delivery can also be directly to any area of the respiratory system (e.g., lungs) via intratracheal intubation. The exact amount of the compositions required will vary from subject to subject, depending on the species, age, weight and general condition of the subject, the severity of the disorder being treated, the particular composition used, its mode of administration and the like. Thus, it is not possible to specify an

exact amount for every composition. However, an appropriate amount can be determined by one of ordinary skill in the art using only routine experimentation given the teachings herein.

[00193] A composition can also be administered by buccal delivery or by sublingual delivery. As used herein "buccal delivery" may refer to a method of administration in which the compound is delivered through the mucosal membranes lining the cheeks. In some embodiment, for a buccal delivery the vaccine composition is placed between the gum and the cheek of a patient. As used herein "sublingual delivery" may refer to a method of administration in which the compound is delivered through the mucosal membrane under the tongue. In some embodiment, for a sublingual delivery the vaccine composition is administered under the tongue of a patient.

[00194] Parenteral administration of the composition, if used, is generally characterized by injection. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions. A more recently revised approach for parenteral administration involves use of a slow release or sustained release system such that a constant dosage is maintained. See, for example, U.S. Pat. No. 3,610,795, which is incorporated by reference herein.

DETAILED DESCRIPTION OF THE INVENTION

[00195] Before the present compounds, compositions, and/or methods are disclosed and described, it is to be understood that this invention is not limited to specific synthetic methods or to specific compositions, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting. Embodiments of the present invention can be freely combined with each other if they are not mutually exclusive.

Pan-Coronavirus Vaccines

[00196] The present invention features preemptive pan-coronavirus vaccines, methods of use, and methods of producing said vaccines, methods of preventing coronavirus infections, etc. The present invention also provides methods of testing said vaccines, e.g., using particular animal models and clinical trials. The vaccine compositions herein can induce efficient and powerful protection against the coronavirus disease or infection, e.g., by inducing the production of antibodies (Abs), CD4+ T helper (Th1) cells, and CD8+ cytotoxic T-cells (CTL).

[00197] The vaccine compositions, e.g., the antigens, herein feature multiple large sequences which may comprise multiple conserved epitopes, that helps provide multiple opportunities for the body to develop an immune response for preventing an infection. Further, the vaccines herein may be designed to be effective against past, current, and future coronavirus outbreaks.

[00198] The vaccine composition comprises multiple large sequences. In certain embodiments, the large sequences are conserved large sequences, e.g., sequences that are highly conserved among human coronaviruses and/or animal coronaviruses (e.g., coronaviruses isolated from animals susceptible to

coronavirus infections).

[00199] The present invention describes the identification of conserved large sequences comprising B cell, CD4+ T cell, and CD8+ T cell epitopes. For example, FIG. 1 shows a schematic of the development of a pre-emptive pan *coronavirus* vaccine featuring multiple conserved large sequences comprising multiple B cell epitopes, multiple conserved CD8+ T cell epitopes, and multiple CD4+ T cell epitopes. The large sequences are derived from sequence analysis of many *coronaviruses*.

[00200] *Coronaviruses* used for determining conserved large sequences may include human SARS-CoVs as well as animal CoVs (e.g., bats, pangolins, civet cats, minks, camels, etc.) as described herein. As an example, FIG. 2A and FIG. 2B show an evolutionary comparison of genome sequences among beta-coronavirus strains isolated from humans and animals. FIG. 2A shows a phylogenetic analysis performed between SARS-CoV-2 strains (obtained from humans (*Homo Sapiens* (black)), along with the animal's SARS-like *Coronaviruses* genome sequence (SL-CoVs) sequences obtained from bats (*Rhinolophus affinis*, *Rhinolophus malayanus* (red)), pangolins (*Manis javanica* (blue)), civet cats (*Paguma larvata* (green)), and camels (*Camelus dromedarius* (Brown)). The included SARS-CoV/MERS-CoV strains are from previous outbreaks (obtained from humans (Urbani, MERS-CoV, OC43, NL63, 229E, HKU1-genotype-B), bats (WIV16, WIV1, YNLF-31C, Rs672, recombinant strains), camel (*Camelus dromedarius*, (KT368891.1, MN514967.1, KF917527.1, NC_028752.1), and civet (Civet007, A022, B039)). The human SARS-CoV-2 genome sequences are represented from six continents. FIG. 2B shows an evolutionary analysis performed among the human-SARS-CoV-2 genome sequences reported from six continents and SARS-CoV-2 genome sequences obtained from bats (*Rhinolophus affinis*, *Rhinolophus malayanus*), and pangolins (*Manis javanica*)).

[00201] Additionally, other *coronaviruses* may be used for determining conserved large sequences (including human SARS-CoVs as well as animal CoVs (e.g., bats, pangolins, civet cats, minks, camels, etc.)) that meet the criteria to be classified as "variants of concern" or "variants of interest." *Coronavirus* variants that appear to meet one or more of the undermentioned criteria may be labeled "variants of interest" or "variants under investigation" pending verification and validation of these properties. In some embodiments, the criteria may include increased transmissibility, increased morbidity, increased mortality, increased risk of "long COVID", ability to evade detection by diagnostic tests, decreased susceptibility to antiviral drugs (if and when such drugs are available), decreased susceptibility to neutralizing antibodies, either therapeutic (e.g., convalescent plasma or monoclonal antibodies) or in laboratory experiments, ability to evade natural immunity (e.g., causing reinfections), ability to infect vaccinated individuals, Increased risk of particular conditions such as multisystem inflammatory syndrome or long-haul COVID or Increased affinity for particular demographic or clinical groups, such as children or immunocompromised individuals. Once validated variants of interest are renamed "variant of concern" by monitoring organizations, such as the CDC.

[00202] The conserved large sequences may be derived from structural (e.g., spike glycoprotein, envelope protein, membrane protein, nucleoprotein) or non-structural proteins of the coronaviruses (e.g.,

any of the 16 NSPs encoded by ORF1a/b).

[00203] In some embodiments, the large sequences are each highly conserved among one or a combination of: SARS-CoV-2 human strains, SL-CoVs isolated from bats, SL-CoVs isolated from pangolin, SL-CoVs isolated from civet cats, and MERS strains isolated from camels. For example, in certain embodiments, the large sequences are each highly conserved among one or a combination of: at least 50,000 SARS-CoV-2 human strains, five SL-CoVs isolated from bats, five SL-CoVs isolated from pangolin, three SL-CoVs isolated from civet cats, and four MERS strains isolated from camels. In certain embodiments, the large sequences are each highly conserved among one or a combination of: at least 80,000 SARS-CoV-2 human strains, five SL-CoVs isolated from bats, five SL-CoVs isolated from pangolin, three SL-CoVs isolated from civet cats, and four MERS strains isolated from camels. In certain embodiments, the large sequences are each highly conserved among one or a combination of: at least 50,000 SARS-CoV-2 human strains in circulation during the COVID-19 pandemic, at least one CoV that caused a previous human outbreak, five SL-CoVs isolated from bats, five SL-CoVs isolated from pangolin, three SL-CoVs isolated from civet cats, and four MERS strains isolated from camels. In certain embodiments, the large sequences are each highly conserved among at least 1 SARS-CoV-2 human strain in current circulation, at least one CoV that has caused a previous human outbreak, at least one SL-CoV isolated from bats, at least one SL-CoV isolated from pangolin, at least one SL-CoV isolated from civet cats, and at least one MERS strain isolated from camels. In certain embodiments, the large sequences are each highly conserved among at least 1,000 SARS-CoV-2 human strains in current circulation, at least two CoVs that has caused a previous human outbreak, at least two SL-CoVs isolated from bats, at least two SL-CoVs isolated from pangolin, at least two SL-CoVs isolated from civet cats, and at least two MERS strains isolated from camels. In certain embodiments, the large sequences are each highly conserved among one or a combination of: at least one SARS-CoV-2 human strain in current circulation, at least one CoV that has caused a previous human outbreak, at least one SL-CoV isolated from bats, at least one SL-CoV isolated from pangolin, at least one SL-CoV isolated from civet cats, and at least one MERS strain isolated from camels. The present invention is not limited to the aforementioned coronavirus strains that may be used to identify conserved large sequences.

[00204] In certain embodiments, one or more of the conserved large sequences are derived from one or more SARS-CoV-2 human strains or variants in current circulation; one or more coronaviruses that has caused a previous human outbreak; one or more coronaviruses isolated from animals selected from a group consisting of bats, pangolins, civet cats, minks, camels, and other animal receptive to coronaviruses; and/or one or more coronaviruses that cause the common cold. SARS-CoV-2 human strains and variants in current circulation may include the original SARS-CoV-2 strain (SARS-CoV-2 isolate Wuhan-Hu-1), and several variants of SARS-CoV-2 including but not limited to Spain strain B.1.177; Australia strain B.1.160, England strain B.1.1.7; South Africa strain B.1.351; Brazil strain P.1; California strain B.1.427/B.1.429; Scotland strain B.1.258; Belgium/Netherlands strain B.1.221; Norway/France strain B.1.367; Norway/Denmark.UK strain B.1.1.277; Sweden strain B.1.1.302; North America, Europe, Asia, Africa, and Australia strain B.1.525; and New York strain B.1.526. The present

invention is not limited to the aforementioned variants of SARS-CoV-2 and encompasses variants identified in the future. The one or more *coronaviruses* that cause the common cold may include but are not limited to strains 229E (alpha coronavirus), NL63 (alpha coronavirus), OC43 (beta coronavirus), HKU1 (beta coronavirus).

[00205] As used herein, the term “conserved” refers to a large sequence that is among the most highly conserved large sequences identified in a sequence alignment and analysis. For example, the conserved large sequences may be the 2 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 3 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 4 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 5 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 6 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 7 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 8 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 9 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 10 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 15 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 20 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 25 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 30 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 40 most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 50 most highly conserved sequences identified. In some embodiments, the conserved sequences may be the 50% most highly conserved large sequences identified. In some embodiments, the conserved large sequences may be the 60% most highly conserved sequences identified. In some embodiments, the large conserved sequences may be the 70% most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 80% most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 90% most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 95% most highly conserved sequences identified. In some embodiments, the conserved large sequences may be the 99% most highly conserved sequences identified. The present invention is not limited to the aforementioned thresholds.

[00206] FIG. 3A shows an example of a systems biology approach utilized in the present invention.

[00207] In some embodiments, the composition comprises one or more large sequences. In some embodiments, the one or more large sequences comprises at least one of one or more conserved *coronavirus* B-cell target epitopes; one or more conserved *coronavirus* CD4⁺ T cell target epitopes; and one or more conserved *coronavirus* CD8⁺ T cell target epitopes

[00208] In other embodiments, the vaccine composition comprises two or more large sequences. In some embodiments, the two or more large sequences comprises at least one of one or more conserved *coronavirus* B-cell target epitopes; one or more conserved *coronavirus* CD4⁺ T cell target epitopes; and one or more conserved *coronavirus* CD8⁺ T cell target epitopes

[00209] In some embodiments, the large sequences comprises one or more conserved *coronavirus* B-cell target epitopes and one or more conserved *coronavirus* CD4⁺ T cell target epitopes. In some embodiments, the large sequences comprises one or more conserved *coronavirus* B-cell target epitopes and one or more conserved *coronavirus* CD8⁺ T cell target epitopes. In some embodiments, the large sequences comprises one or more conserved *coronavirus* CD8⁺ target epitopes and one or more conserved *coronavirus* CD4⁺ T cell target epitopes. In some embodiments, the large sequences comprises one or more conserved *coronavirus* CD8⁺ target epitopes. In some embodiments, the large sequences comprises one or more conserved *coronavirus* CD4⁺ target epitopes. In some embodiments, the large sequences comprises one or more conserved *coronavirus* B cell target epitopes.

[00210] In some embodiments, the vaccine composition comprises one or more conserved *coronavirus* CD8⁺ target epitopes. In some embodiments, the vaccine composition comprises one or more conserved *coronavirus* CD4⁺ target epitopes. In some embodiments, the vaccine composition comprises one or more conserved *coronavirus* B cell target epitopes.

[00211] As will be discussed herein, in certain embodiments, the vaccine composition comprises whole spike protein, one or more coronavirus CD4⁺ T cell target epitopes; and one or more coronavirus CD8⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises at least a portion of the spike protein (e.g., wherein the portion comprises a trimerized SARS-CoV-2 receptor-binding domain (RBD)), one or more coronavirus CD4⁺ T cell target epitopes; and one or more coronavirus CD8⁺ T cell target epitopes. In some embodiments, the one or more coronavirus CD4⁺ T cell target epitopes; and one or more coronavirus CD8⁺ T cell target epitopes may be in the form of a large sequence.

[00212] The large sequences may be each separated by a linker. In certain embodiments, the linker allows for an enzyme to cleave between the large sequences. The present invention is not limited to particular linkers or particular lengths of linkers. As an example, in certain embodiments, one or more large sequences may be separated by a linker 2 amino acids in length. In certain embodiments, one or more large sequences may be separated by a linker 3 amino acids in length. In certain embodiments, one or more large sequences may be separated by a linker 4 amino acids in length. In certain embodiments, one or more large sequences may be separated by a linker 5 amino acids in length. In certain embodiments, one or more large sequences may be separated by a linker 6 amino acids in length. In certain embodiments, one or more large sequences may be separated by a linker 7 amino acids in length. In certain embodiments, one or more large sequences may be separated by a linker 8 amino acids in length. In certain embodiments, one or more large sequences may be separated by a linker 9 amino acids in length. In certain embodiments, one or more large sequences may be separated by a linker 10 amino acids in length. In certain embodiments, one or more large sequences may be separated

by a linker from 2 to 10 amino acids in length.

[00213] Linkers are well known to one of ordinary skill in the art. Non-limiting examples of linkers include AAY, KK, and GPGPG.

[00214] The large sequences may be derived from structural proteins, non-structural proteins, or a combination thereof. For example, structural proteins may include spike proteins (S), envelope proteins (E), membrane proteins (M), or nucleoproteins (N).

[00215] In some embodiments, the large sequences are derived from at least one SARS-CoV-2 protein. The SARS-CoV-2 proteins may include ORF1ab protein, Spike glycoprotein, ORF3a protein, Envelope protein, Membrane glycoprotein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, Nucleocapsid protein, and ORF10 protein. The ORF1ab protein provides nonstructural proteins (Nsp) such as Nsp1, Nsp2, Nsp3 (Papain-like protease), Nsp4, Nsp5 (3C-like protease), Nsp6, Nsp7, Nsp8, Nsp9, Nsp10, Nsp11, Nsp12 (RNA polymerase), Nsp13 (5' RNA triphosphatase enzyme), Nsp14 (guanosineN7-methyltransferase), Nsp15 (endoribonuclease), and Nsp16 (2'-O-ribose-methyltransferase).

[00216] The SARS-CoV-2 has a genome length of 29,903 base pairs (bps) ssRNA (SEQ ID NO: 1). Generally, the region between 266-21555 bps codes for ORF1ab polypeptide; the region between 21563-25384 bps codes for one of the structural proteins (spike protein or surface glycoprotein); the region between 25393-26220 bps codes for the ORF3a gene; the region between 26245-26472 bps codes for the envelope protein; the region between 26523-27191 codes for the membrane glycoprotein (or membrane protein); the region between 27202-27387 bps codes for the ORF6 gene; the region between 27394-27759 bps codes for the ORF7a gene; the region between 27894-28259 bps codes for the ORF8 gene; the region between 28274-29533 bps codes for the nucleocapsid phosphoprotein (or the nucleocapsid protein); and the region between 29558-29674 bps codes for the ORF10 gene.

[00217] The large sequences may comprise a T-cell epitope restricted to a large number of human class 1 and class 2 HLA haplotypes and not restricted to HLA-0201 for class 1 or HLA-DR for class 2. The conserved large sequences may be restricted to human HLA class 1 and 2 haplotypes. In some embodiments, the conserved epitopes are restricted to cat and dog MHC class 1 and 2 haplotypes.

Large Sequences

[00218] The antigen may comprise large sequences, such as conserved large sequences that are highly conserved among human and animal coronaviruses. As used herein, the term large sequence refers to a sequence having at least 25 amino acids or at least 75 nucleotides. The large sequences comprise epitopes, such as the conserved epitopes described herein.

[00219] In order to identify the conserved large sequences, sequence alignments and analysis were performed as described herein as well as below.

[00220] *Sequence comparison among SARS-CoV-2 and previous coronavirus strains:* Sequence homology analysis we performed and compare the Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) isolate Wuhan-Hu-1, to complete genome with sequences of SARS-CoV-2 variants, common cold corona virus strains (HKU1 genotype B, CoV-OC43, CoV-NL63, and CoV-229E), SARS-CoV-Urbani, MERS and coronavirus strains from bats (*Rhinolophus affinis* and *R. malayanus*), pangolin (*Manis javanica*), civet cats (*Paguma larvata*), and camel (*Camelus dromedarius* and *C. bactrianus*).

[00221] The human SARS-CoV-2 variant genome sequences were retrieved from the GISAID database, representing major Variants of Concern which are known for their high degree of transmissibility and pathogenicity. The sequences used in this study are 20A.EU1 from Spain (EPI_ISL_691726-hCoV-19-VOC-20A.EU1), 20A.EU2 from Australia (EPI_ISL_418799-hCoV-19-VOC-20A.EU2), B.1.1.7 from England (EPI_ISL_581117-hCoV-19-VOC-B.1.1.7), B.1.351 from South Africa (EPI_ISL_660615-hCoV-19-VOC-B.1.351), P.1 from Brazil (EPI_ISL_581117-hCoV-19-VOC-P.1), CAL.20C from California (EPI_ISL_730092-hCoV-19-VOC-B.1.427/B.1.429), B.1.258 from Scotland (EPI_ISL_858559-hCoV-19-VOC-B.1.258), B.1.221 from Belgium/Netherlands (EPI_ISL_734790-hCoV-19-VOC-B.1.221), B.1.367 from Norway/France (EPI_ISL_541518-hCoV-19-VOC-B.1.367), B.1.1.277 from Netherlands/Denmark/UK (EPI_ISL_500783-hCoV-19-VOC-B.1.1.277), B.1.1.302 from Sweden (EPI_ISL_717929-hCoV-19-VOC-B.1.1.302). Similarly, HKU1 genotype B (AY884001), CoV-OC43 (KF923903), CoV-NL63 (NC_005831), and CoV-229E (KY983587), SARS-CoV-Urbani (AY278741.1), MERS (NC_019843).

[00222] Bat CoV strains used in this analysis include strains RaTG13 (MN996532.2), Rs672/2006 (FJ588686.1), YNLF_31C (KP886808.1), WIV1 (KF367457.1), WIV16 (KT444582.1), ZXC21 (MG772934.1), RmYN02 (EPI_ISL_412977), bat-RmYN01 (EPI_ISL_412976), MERS-Bat-CoV/P.khulii/Italy/206645-63/2011 (MG596803.1). More-so, five genome sequences representing Pangolin (MT040333.1-PCoV_GX-P4L, MT040334.1-PCoV_GX-P1E, MT040335.1-PCoV_GX-P5L, MT040336.1-PCoV_GX-P5E, MT072864.1-PCoV_GX-P2V, MT121216.1-PCoV-MP789), three Civet cat specific genome sequences (AY572034.1, AY686864.1, AY686863.1), and four CoV sequences from camels (NC_028752.1, KF917527.1, MN514967.1, KT368891.1) were included in this sequence homology analysis aimed at evaluating the most conserved regions in different structural and non-structural proteins in CoV genome. These sequences were obtained either from National Center for Biotechnology Information (NCBI) or Global initiative on sharing all influenza data (GISAID). For phylogenetic analyses, SARS-CoV-2 full-genome sequences were aligned with CLUSTAL W using MEGAX. All the SARS-CoV-2 sequences were compared to existing genomes using online NCBI BLAST.

[00223] *Determination of SARS-CoV-2 Sequence Conservation:* Each Wuhan-Hu-1 (GeneBank: NC_045512.2) specific structural (Spike glycoprotein (YP_009724390.1), Membrane protein

(YP_009724393.1), Envelope protein (YP_009724392.1), Nucleocapsid phosphoprotein (YP_009724397.2)), and non-structural proteins (ORF1a/b polyprotein (YP_009724389.1), ORF3a (YP_009724391.1), ORF6 (YP_009724394.1), ORF7a (YP_009724395.1), ORF7b (YP_009725318.1), ORF8 (YP_009724396.1), and ORF10 (YP_009725255.1)) protein sequences were compared against the consensus protein sequences from SARSCoV and MERS-CoV and the protein sequences from closest relative cross species CoV strains using the Nucleotide BLAST (blastn) algorithm to compute the pairwise identity between Wuhan-Hu-1 proteins and their comparison target.

[00224] Further as the present invention is interested in the highly similar sequences across CoV strains, megablast was performed. For each of the queried sequences, Query coverage, E value, Percent identity were determined. The queried homology obtained against one bat CoV strain RmYN01, which was found earlier to be phylogenetically less similar to SARS-CoV-2, but has more genetic similarities with SARS-CoV-Urbani was taken as a standard to ascertain the homologous sequences across CoV strains. The strategy was helpful to find out how genetically more conserved regions among different CoVs. This sequence has a query coverage of 59%, and a percent identity of 78.73% when compared against the SARS-CoV-2 genome sequence. It has five matched regions which further showed sequence homology among other CoVs as well. Matched region 1 spanned between 1bp-1580bp (fragment) showed sequence homology with nsp1 (leader protein), nsp2, and nsp3, whereas matched region 2 spanned between 3547bp-7096bp (fragment 2) showed sequence homology with multiple subunits of ORF1a/b like 3CLpro, nsp6, nsp7, nsp8, nsp9, nsp10, RNA dependent RNA polymerase, helicase, nsp14, nsp15, and nsp16. Interestingly, a major region spanning in the non-annotated region of the ORF1a/b between 17472bp-21156bp (fragment 3) also showed sequence identity. The fourth stretch of sequence identity spanned through 22584bp-24682bp (fragment 4) covering a section of the Spike glycoprotein, that importantly covers the major Receptor Binding Domain in the SARS-CoV-2 as well. The last segment of the homologous sequence showed percent identity with regions specific to the ORF3a, Envelope protein, Membrane protein, ORF6, and ORF7a (26193bp- 27421bp; fragment 5).

[00225] In some embodiments, five fragments from the SARS-CoV-2 Wuhan Strain were found to be highly conserved (1bp- 1580bp (fragment 1), 3547bp- 12830bp (fragment 2), 17472bp- 21156bp (fragment 3), 22584bp- 24682bp (fragment 4), and 26193bp- 27421bp (fragment 5). Next, each fragment underwent another round of sequence homology analysis.

[00226] In some embodiments, the vaccine composition comprises one large sequence. In some embodiments, the vaccine composition comprises one or more large sequences. In some embodiments, the vaccine composition comprises two or more large sequences. In some embodiments, the vaccine composition comprises three or more large sequences. In some embodiments, the vaccine composition comprises four or more large sequences. In some embodiments, the vaccine composition comprises five or more large sequences, e.g., 5, 6, 7, 8, etc.

[00227] In some embodiments, the large sequences are derived from a whole protein sequence expressed by SARS-CoV-2. In other embodiments, large sequences are derived from a partial protein

sequence expressed by SARS-CoV-2. In some embodiments, the large sequence of said proteins comprise B cell epitopes and T-cell epitopes that are restricted to a large number, e.g., from 3 to 10, different haplotypes that encompass 100% of the population regardless of race and ethnicity) of human class 1 and class 2 HLA haplotypes, so they are not restricted only to HLA-0201 for class 1 or HLA-DR1 for class 2.

[00228] As previously discussed, the large sequences may be highly conserved among human and animal coronaviruses. In some embodiments, the large sequences are derived from one or a combination of: one or more SARS-CoV-2 human strains or variants in current circulation; one or more *coronaviruses* that has caused a previous human outbreak; one or more *coronaviruses* isolated from animals selected from a group consisting of bats, pangolins, civet cats, minks, camels, and other animal receptive to coronaviruses; and/or one or more *coronaviruses* that cause the common cold.

[00229] As previously discussed, the SARS-CoV-2 human strains or variants in current circulation may include strain B.1.177; strain B.1.160, strain B.1.1.7; strain B.1.351; strain P.1; strain B.1.427/B.1.429; strain B.1.258; strain B.1.221; strain B.1.367; strain B.1.1.277; strain B.1.1.302; strain B.1.525; strain B.1.526, strain S:677H, and strain S:677P. The *coronaviruses* that cause the common cold may be selected from: 229E alpha coronavirus, NL63 alpha coronavirus, OC43 beta coronavirus, and HKU1 beta coronavirus.

[00230] The large sequence(s) may be derived from structural proteins, non-structural proteins, or a combination thereof. The large sequence(s) may be selected from ORF1ab protein, Spike glycoprotein (e.g., the RBD), ORF3a protein, Envelope protein, Membrane glycoprotein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, Nucleocapsid protein, and/or an ORF10 protein. Note the ORF1ab protein comprises nonstructural protein (Nsp) 1, Nsp2, Nsp3, Nsp4, Nsp5, Nsp6, Nsp7, Nsp8, Nsp9, Nsp10, Nsp11, Nsp12, Nsp13, Nsp14, Nsp15 and Nsp16.

[00231] In some embodiments, a large sequence comprises conserved fragments from over 150,000 CoV strains circulating in the majority of countries around the world (Table 1, FIG. 4). In some embodiments, fragment 1 comprises the base pairs 1-1580. In some embodiments, fragment 1 may comprise the proteins Nsp1, Nsp2, and Nsp3 as well as unannotated regions (FIG. 5). In some embodiments, fragment 2, comprises the base pairs 3547-12830. In some embodiments, fragment 2 may comprise the proteins Nsp5, Nsp6, Nsp7, Nsp8, Nsp9, Nsp10, Nsp11, Nsp12, Nsp13, Nsp14, Nsp15, Nsp16, as well as unannotated regions (FIG. 6). In some embodiments, fragment 3 comprises the base pairs 17472-21156. In some embodiments, fragment 3 comprises unannotated regions (FIG. 7). In some embodiments, fragment 4 comprises the base pairs 22584-24682. In some embodiment, fragment 4 comprises the spike glycoprotein (FIG. 8). In some embodiments, fragment 5 comprises the base pairs 26193-27421. In some embodiments, fragment 5 comprises the proteins ORF3a, Envelope (E), Membrane (M), ORF6, ORF7a, as well as unannotated regions (FIG. 9).

[00232] Table 1:

Fragment No:	Proteins	Sequence	SEQ ID NO:
1 (1bp-1580bp)	Nsp1, Nsp2, Nsp3	GACGTGCTAGTACGTGGCTTCGGGGACTCTGTGGAAGAGGCCCTAT CGGAGGCACGTGAACATCTTAAAAGTGGCACTTGTGGCATAGTAGAG CTGGAAAAAGGCGTATTGCCTCAGCTTGAACAGCCCTATGTGTTCAAT AAACGATCTGACGCCAGGGCACTGGTCATGGCCACAAGGTCTGTG AGCTAGTTGCTGAATTGGATGGCGTTCAAGTTCGGTCGTAGCGGTATA ACACTGGGAGTACTCGTGCCACACGTTGGCGAGACCCCAATTGCATA CCGCACTGTTCTTCTTCGTAAGAATGGTAATAAGGGAGCCGGTGGCC ATAGCTTTGGCATCGATCTAAAGTCATATGACTTAGGTGACGAGCTTG GCACTGATCCCATTGAAGATTATGAACAAAACCTGGAACACTAAACATG GCAGTGGTGCCCTTCGTGAACCTCACTCGTGAGCTCAATGGAGGAGT AGTTACTCGCTATGTCGACAACAATTTCTGTGGCCAGATGGCTACCC CCTTGAATGCATTAAAGACCTTCTCGCTCGTGCGGGCAAGTCAATGT GCACTCTTTCTGAACAACCTTGATTTTATCGAGTCGAAGAGAGGTGTCT ACTGCTGTGCTGAACATGAGCATGAAATTGCTTGGTTTACCGAACGC TCTGAAAAGAGCTATGAGCACCAGACACCCCTTCGAGATCAAGAGTGC CAAGAAATTTGACACTTTCAAAGGGGAATGCCCAAAGTTTGTATTTCC TCTCAATTCTAGAGTCAAAGTCATTCAACCACGTGTTGAAAAGAAAAA GACTGAAGGTTTCATGGGGCGTATACGCTCTGTGTACCCTGTTGCAT CCCCTGGGGATTGTAACGATATGCACTTGTCTACCTTGATGAAATGTA ATCATTGTGATGAAGTTTCATGGCAGACGTGCGACTTTCTCAAAGCCA CTTGTGAACAATGTGGCACTGAAAACCTTAGTCTGTGAAGGACCCACT ACATGTGGATACCTACCTACTAATGCTGTACTTAAATGCCTTGTCTTG CTTGTCAAGATCCAGAGATTGGACCTGAGCATAGTGTTCAGACTATC ACAACCCTCAAAACATTGAAACTCGACTCCGCAAGGGAGGTAGGACT AAATGTTTTGGTGGGTGTGTGTTTGCCTACGTTGGCTGCTATAACAAG CGTGCCCTACTGGGTTCCCTCGTGTAGTGCCGATATTGGTGCAAACCA TACTGGCATTACTGGAGACAATGTGGAGACTTTAAATGAAGATCTCCT GGAGATACTGCATCGTGAACGTGTTAATGTTAACATTGTTGGCGATTT TCAGTTGAATGAAGAGGTTGCTATTATTCTAGCATCTTTCTCTGCTTCT ACTAGTGCCCTTTATTGACACTGTAAAGGGCCTTGACTACAAGACCTTC AAAGCCATTGTTGAATCCTGTGGAACTACAAAGTTACCAAAGGAAAA CCTGTCCAAGGAGCTTGGAACATTGGCCAGCAAAAATCTATTTTGACA CCGCTGTGTGTTTTCCATCACAGGCTGCCAGTGTCATTAGATCAATC TTTTCTCGCAC	182
2 (2547bp-12830bp)	Nsp5, Nsp6, Nsp7, Nsp8, Nsp9, Nsp10, Nsp11, Nsp12, Nsp13, Nsp14, Nsp15, Nsp16	AAAATTAAGGCTTGCATCGAAGAGGTCACTACAACACTGGAAGAGAC TAAGTTTCTTACCAATAAGTTGCTTCTTTTCTGATATCAGCGGTAAA CTTTACCAAGATTCTCAGAAATATGCTTAGAGGTGAGGACGTGTCTTTC CTTGAGAGAGATGCGCCTTACATGGTAGGTGATGTTATCAATAGTGGT GATATTACCTGCGTTGTAATACCTTCTAAGAAGGCTGGTGGTACTACA GAAATGCTTGCAAGAGCATTGAAGAAAGTGCCARTTGATGAGTATATA ACCACATAYCCTGGWCAAGGWTGTGCTGGTTATACACTTGAKGAAGC TARGACTGCTCTTAARAARTGCAAATCTGCAYTKTAYGKTACCTTCA GAATCACCTAATGCTAAGGAAGAGATTCTAGGAACCGTATCTTGGAAAT TTGAGAGAAAATGCTTGCTCAGCTGAAGAGACAAGAAAATTAATGCCT ATCTGCATGGATGTCAGAGCCATAATGGCCACCATCCAACGCAAGTA CAAAGGAATTAAAATTCAAGAAGGCATCGTTGACTATGGTGTCCGATT CTTCTTTTATACTAGTAAAGAGCCTGTAGCTTCTATTATTACGAAGCTG AACTCTCTAAATGAGCCACTTGTCAATGCCAATTGGTTATGTGACA CATGGTTTTAATCTTGAAGAGGCTGCGCGCTGTATGCGTTCTCTTAAA GCTCCTGCCGTAGTGTGATCATCACCAGATGCTGTTACTACATAT AATGGATACCTCACTTCGTATCAAAGACATCTGAGGAGCACTTTGTG	183

		GAAACAGTTTCTTTGGCTGGCTCTTACAGAGATTGGTCCTATTCAGGA CAGCGTACAGAGTTAGGTGTTGAATTTCTTAAGCGTGGTGACAAAATT GTGTACCACACTTTGGAGAGCCCCGTCAAGTTCCATCTTGACGGTGA GGTTCTTCCACTTGACAAATTAAAGAGTCTTTATCCCTACGGGAGGT TAAGACTATAAAAGTGTTCACAACTGTGGACAATACTAATCTCCACACA CATCTTGTGGATATGTCTATGACATATGGACAGCAGTTTGGTCCAACAT ATTTGGATGGTGCTGATGTTACAAAAATTAAACCTCATGTAAATCATGA GGGTAAGACTTTCTTTGTATTACCTAGTGATGACACACTACGTAGTGA AGCTTTTGAGTACTACCACGCTCTTGATGAGAGTTTCCTTGGTAGATA CATGTCTGCTTTAAACCACACAAAGAAATGGAAATCCCTCAAGTTGG TGGTTTGACTTCCATTAAGTGGGCTGATAACAATTGTATTTGTCTAGT GTTTTATTAGCACTTCAACAAATTGAAGTTAAATTTAATGCCCCAGCAC TACAAGAAGCTTACTATAGAGCTCGTGCTGGTGATGCTGCTAATTTTT GTGCACTTATACTCGCTTACAGTAATAAAAACTGTTGGCGAGCTGGGTG ATGTCAGAGAAACTATGGCCCATCTTTACAGCATGCTAATTTGGAATC TGCAAAGCGAGTTCTTAATGTGGTGTGTAAACATTGCGGCCAGAAAA CTACTACCTTAACGGGTGTAGAGGCTGTGATGTACATGGGTACTCTGT CTTATGATAATCTTAAGACAGGTGTTTCTGTTCCATGTGTGTGGTC GTGACGCTACACAATATTTAGTACAACAAGAGTCTTCTTTTGTATGAT GTCCGCACCACCTGCTGAATATAAATTACAGCAAGGTACATTCTTATGT GCAAATGAATACACTGGTAATTATCAGTGTGGTCATTACACTCATATAA CTGCTAAGGAGACCCCTATCGTATTGATGGAGCTCACCTTACAAAGA TGTCAGAGTATAAAGGGCCAGTGACTGATGTGTTCTACAAGGAAACAT CTTACACTACAACCATCAAGCCTGTGTCATATAAACTCGATGGAGTTAC TTACACAGAGATTGAACCAAATTTGGATGGGTATTATAAAAAGGATAAT GCTTACTATACGGAGCAGCCTATAGACCTTGTACCAACTCAACCACTA CCAAATGCGAGTTTTGATAATTTCAAACCTCACATGTTCTAATATAAAATT CGCTGATGATTTAAATCAAATGACAGGCTTCACAAAGCCAGCTTCACG AGAGCTATCTGTCACATTCTTCCAGACTTGAATGGCGATGTAGTGGC TATTGACTATAGACACTACTCAGCGAGTTTCAAGAAAGGTGCTAGATTA CTGCATAAGCCAATTGTTTGGCATATCAATCAGGCTACAACCAAGACA ACGTTTACAGACCAACACTTGGTGTTTACGTTGTCTTTGGAGTACAAAA CCAGTAGATACTTCAAATTCATTTGAAGTTCTGGCAGTAGAAGACACA CAAGGAATGGACAATCTTGCTTGTGAAAGTCAAAGACCCACCTCTGA AGAAGTAGTGGAAAAATCCTACCATACAGAAGGAAGTCATAGAGTGTGA CGTGAAAACTACCGAAGTTGTAGGCAATGTCATACTTAAACCATCAGA TGAAGGTGTTAAAGTAACACAAGAGTTAGGGCATGAGGATCTTATGGC TGCCTATGTGGAAAAATACAAGCATTACCATTAAAGAAACCTAATGAGCTT TCATTAGCCTTAGGTTTAAAAACAATTGCCACTCATGGTATTGCTGCAA TTAACAGTGTTCCGTGGAGTAAATTTTGGCTTATGTCAGACCATTC TAGGACGAACAGCAATCACACATCAAACCTGTGCTAAGAGATTAGTAC AGCGTGATTTTAAACAACTACATGCCCTATGTGCTTACATTATTGTTCCA ATTGTGTACTTTTACCAAAAGTACAAATCTAGAATTAGAGCTTCACTA CCTACGACTATTGCTAAAAATAGTGTTAAGGGTGTGCTAAATTATGTT TGGATGCTGGCATCAATTATGTAAAGTCACCCAAATTTTCTAAATTGTT CACTATTGCAATGTGGCTATTATTGTTAAGCATTGCTTAGGTTCACTA ATCTATGTAACGACGCTTTAGGTGTATTATTGTCCAACCTTTGGAGCTC CTTCCTATTGTAGTGGCGTTAGAGAATCGTATCTCAATTCCTCTAATGT TACTACTATGGACTTCTGTGAAGGTTCTTTTCCTTGCAGCGTTTGTTTA AGTGGATTAGACTCGCTTGATTCCATCCAGCTCTTGAAACCATACAG GTAACGATTTTCATCGTATAAGCTAGACTTGACAATTTTAGGTCTGGCTG CTGAGTGGTTTTTGGCATATATGTTGTTCAAAAATTTCTTTTATTATTA GGTCTTTCAGCTATAATGCAGGTGTTCTTTGGCTATTTTGTAGTCATT TCATCAGCAATTCTTGGCTTATGTGGTTTATCATTAGTATCGTACAAATG GCACCCGTTTCCGCAATGGTTAGGATGTACATTTTCGTTGCTTCTTTC TACTACATATGGAAGAGCTATGTTTCATATTATGGATGGTTGTACTTCATC TACTTGCAATGATGTGCT	
--	--	--	--

4 (22584bp- 24683bp)	spike glycopro tein	TACCAAGCTACTAGAGTAGTGGTACTTTTCATTTGAGCTTCTAAATGCAC CTGCCACAGTGTGTGGACCAAAATTGTCCACATCACTAATTAAGAACC AGTGTGTCAATTTTAATTTCAATGGACTCAAGGGTACTGGTGTGTTGA CTGACTCGTCCAAAAAGTTTCAGTCTTTTCAACAATTTGGAAGGGATG CATCTGATTTTACTGACTCAGTACGCGACCCTCAGACACTTCAAAATAC TTGACATTTACCATGTTTCATTTGGTGGTGTGAGTGAATAACACCAG GAACAAATGCTTCATCTGAAGTAGCCGTTCTATACCAAGATGTAAACT GCACTGATGTTCCACGCGCCATACGTGCTGACCAACTCACACCTGCT TGGCGTGTCTTACTCTGCTGGAGTAAATGTGTTTCAAACCTCAGGCTGG CTGTTTAATAGGAGCGGAACATGTCAATGCTTCATATGAGTGTGACAT TCCCATTGGTGCAGGCATTTGTGCTAGTTACCATACAGCTTCCCTTTT ACGTAATACAGGCCAGAAATCAATTGTGGCCTATACTATGTCACCTGG TGCTGAAAACCTCAATTGCTTATGCTAATAACTCAATTGCCATACCTACA AATTTTTCAATCAGTGTCACAACTGAAGTGATGCCTGTTTCAATGGCT AAGACATCAGTAGATTGTACAATGTACATCTGTGGTGACTCTCAGGAG TGCAGCAACTTACTACTTCAGTATGGTAGCTTTTGCACACAATTAATC GTGCCCTTTCAGGCATTGCTGTTGAACAGGACAAAAACACTCAAGAG GTTTTTGCCCAAGTTAAACAAATGTATAAGACACCAGCCATAAAAGATT TTGGTGGCTTTAATTTCTCACAATATTGCCTGACCCTTCTAAGCCAAAC AAAAAGATCATTTATTGAGGATTTACTCTTCAACAAAGTGACTCTCGCT GATGCTGGCTTTATGAAGCAATACGGCGAATGCCTAGGCGATATTAGT GCTAGAGATCTCATTTGTGCGCAGAAGTTCAATGGACTCACTGTCCTT CCACCTCTACTCACGGATGAAATGATTGCTGCTTACACCGCCGCTCTT GTCAGCGGTACTGCTACTGCTGGTTGGACATTTGGTGCAGGTGCTGC TCTACAAATACCTTTTGTCTATGCAAATGGCTTATAGGTTCAATGGCATT GGAGTTACTCAAATGTTCTCTATGAGAACCAGAAGCAGATCGCTAAC CAATTTAACAAGGCGATCAGTCAAATCAAGAATCACTTACTACTACTT CAACTGCATTGGGCAAGCTGCAAGACGTCGTCAACCAGAATGCTCAA GCATTGAACACACTTGTAAACAATAAGTTCTAACTTTGGTGCAATTT CAAGTGTTTTAAATGACATTCTGTCTCGACTYGACAAAGTTGAGGCTG AAGTGCAAATTTGATAGGTTGATTACTGGCAGATTACAAAGCCTTCAGA CCTATGTAACACAACAATAATCAGAGCTGCTGAAATCAGAGCTTCTG CCAATCTTGCTGCCACTAAGATGTCCGAGTGTGTTCTTGGACAATCAA AAAGAGTTGACTTTTGTGGAAAAGGCTATCATCTTATGTCTTCCCTC AAGCAGCCCCACATGGTGTCTGCTTCTTACATGTCACATACGTGCCAT CGCAAGAAAGAACTTCACCACTGCCCCAGCAATCTGCCATCAAGGC AAGGCACACTTCCCTCGTGAAGGTGTTTTTGTATCTAATGGCACTTCT TGGTTTATCACACAGAGGAACCTTTTACCACAAAAATTAATACACAG ACAATACATTTGTCTCTGGAATTTGTGATGTCGTTATTGGCATCATCAA CAATACTGTTTATGATCCTCTGCAACCTGAGCTTGACTCATTTAAAGAA GAGCTGGACAAGTACTTCAAAAACACACGTACCTGATGTRGATCT TGGCGACATCTCAGGCATTAATGCTTCAGTCGTCAATATTCAAAAAGA AATTGACCGCCTCAATGAGGTTGCCAAAAATCTAAATGAATCGCTCAT CGATCTTCAAGAACTTGGAAAATATGAGCA	184
5 (26193bp- 27421bp)	ORF3a, Envelope (E), Membrane (M), ORF6, ORF7a	CAGTAACACTTGCTTGCTTTGTGCTTGCTGCTGTTTACAGAATTAATT GGGTGACTGGCGGAATTGCRATTGCAATGGCTTGATTGTAGGCTTG ATGTGGCTTAGCTACTTCRTTGCTTCTTTCAGGCTGTTTGC GCGCACC CGCTCWATGTGGTCATTCAACCCAGAAACYAACATTCTTCTCAATGTG CCTCTTCGRGGRACAATYTTGACCAGACCGCTCATGGARAGTGAAC TGTCAATGGTGCTGTGATCATTCGTGGTCACCTGCGAATGGCTGGAC ACTCYCTWGGGCGCTGTGACATTAAGGACCTGCCAAAAGAGATCACT GTGGCTACATCACGAACGCTTTCTTATTACAAATTAGGAGCTTCGCAG CGTGTAGGCACTGACTCAGGTTTTGCTGCATACAACCGCTACCGTATT GGAACTACAAATTAATACAGACCACGCCGGTAGCAACGACAATATT GCTTTGCTAGTACAGTAAGTGACAACAGATGTTTCATCTAGTTGACTT CCAGGTTACAATAGCGGAGATATTGATTATCATTATGAGGACTTTCAGG ATTGCCATCTGGAATCTTGATGTAATAATAAGTTCAATAGTGAGACAAT	185

		TATTTAAGCCTCTAACTAAGAAGAATTATTCTGAGTTAGATGATGAAGA ACYTATGGAGATTGATTATCCATAAAACGAACATGAAAATTATCCTCTTC CTGACTTTGATTTCACTTGCATTTTGTGAGTTATATCATTATCAGGAGT GTGTTAGAGGTACAACCTGTACTATTTAAAGAACCCTTGCCCATCRGGAA CGTACGAGGGCAATTCACCATTTCCACCCTCTTGCTGACAACAAATTTG CACTAAGTTCATTAGCACACATTTTGCTTTTGCTTGCTGACGGTA CTCGACATACCTATCAGCTTCGTGCAAGATCAGTTTCTCCAAAACCTCT TCATCAGGCAAGAGGAATTTCAATCAAGAGCTCTATTCACCACTTTTTC TCATTGTTGCCGCTCTAGTATTTATAATACTTTGCTTCACCATTAAGAGA AAGACCGAATGAGTGAGCTCACTTTAATTGACTTCTATTTGTGCTTTT AGCCTTTCTGCTATTCCTTGTTTTAATAATGCTCATCATATTTTGTTCT CCTTGAGATTCAAGATTCTGAAGAGCCATGTCCAAAAGTCTAAACGA ACATGAAACTTCTCATTGTTTT	
--	--	--	--

[00233] In some embodiments, the large sequences are not limited to the above mentioned conserved fragments.

[00234] In certain embodiments, the large sequence comprises spike glycoprotein (S) or a portion thereof (e.g., the RBD), nucleoprotein or a portion thereof, membrane protein or a portion thereof, and/or ORF1a/b or a portion thereof (see Table 9, SEQ ID NO: 139). In certain embodiments, the large sequence comprises Spike glycoprotein (S) or a portion thereof (e.g., the RBD), Nucleoprotein or a portion thereof, and ORF1a/b or a portion thereof. In further embodiments, the large sequence comprises Spike glycoprotein (S) or a portion thereof (e.g., the RBD), and Nucleocapsid protein or a portion thereof (see Table. 9, SEQ ID NO: 140).

[00235] As will be discussed herein, in certain embodiments, the vaccine composition comprises whole spike protein, one or more coronavirus CD4+ T cell target epitopes; and one or more coronavirus CD8+ T cell target epitopes. In certain embodiments, the vaccine composition comprises at least a portion of the spike protein (e.g., wherein the portion comprises a trimerized SARS-CoV-2 receptor-binding domain (RBD)), one or more coronavirus CD4+ T cell target epitopes; and one or more coronavirus CD8+ T cell target epitopes. In some embodiments, the one or more coronavirus CD4+ T cell target epitopes; and one or more coronavirus CD8+ T cell target epitopes are in the form of a large sequence.

[00236] In some embodiments, the large sequence(s) are derived from a full-length spike glycoprotein. In other embodiments, the large sequence(s) are derived from a portion of the spike glycoprotein. In some embodiments, the transmembrane anchor of the spike protein has an intact S1–S2 cleavage site. In some embodiments, the spike protein is in its stabilized conformation. In some embodiments, the spike protein is stabilized with proline substitutions at amino acid positions 986 and 987 at the top of the central helix in the S2 subunit. In some embodiments, the composition comprises a SARS-CoV-2 receptor–binding domain (RBD). In some embodiments, the composition comprises a trimerized SARS-CoV-2 receptor–binding domain (RBD). In some embodiments, the trimerized SARS-CoV-2 receptor–binding domain (RBD) sequence is modified by the addition of a T4 fibrin–derived foldon trimerization domain. In some embodiments, the addition of a T4 fibrin–derived foldon trimerization domain increases immunogenicity by multivalent display.

[00237] In some embodiments, the spike protein comprises Tyr-489 and Asn-487 (e.g., Tyr-489 and Asn-487 help with interaction with Tyr 83 and Gln-24 on ACE-2). In some embodiments, the spike protein comprises Gln-493 (e.g., Gln-493 helps with interaction with Glu-35 and Lys-31 on ACE-2). In some embodiments, the spike protein comprises Tyr-505 (e.g., Tyr-505 helps with interaction with Glu-37 and Arg-393 on ACE-2). In some embodiments, the composition comprises a mutation 682-RRAR-685 → 682-QAAQ-685 in the S1-S2 cleavage site.

[00238] In some embodiments, the spike protein comprising the large sequence(s) comprises at least one proline substitution. In some embodiments, the spike protein comprising the large sequence(s) comprises at least two proline substitutions. For example, the proline substitution may be at position K986 and V987.

[00239] Non-limiting examples of sequences are disclosed in Table 2.

Table 2:

	Sequence:	SEQ ID NO:
SARS-CoV-like Spike-S1-NTD 13bp-304bp	SQCVNLTTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFF SNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIF GTTLDSTQSLIVNNATNVVIVKCEFQFCNDPFLGVYHKNKSW MESEFRVYSSANNCTFEYVSQPFMDLEGKQGNFKNLREFVFNID GYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRS YLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCAL DPLSETKCTLK	186
SARS-CoV-2 Spike-S1-RBD 319bp-541bp	RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRKISNCVADY SVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQIAP GQTGKIADYNYKLDDFTGCVIAWNSNNLDSKVGGNYYLYRLFRK SNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVG YQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNF	187
CoV Spike S1-S2_S2 543bp-1,208bp	FNGLTGTGVLTESNKKFLPFQFGRDIADTTDAVTRAGCLIGAEHVN NSYECDIPIGAGICASYQTQTNRDPQTLEILDITPCSFGGVSITPGT NTSNQVAVLYQDVNCTEVPVAIHADQLTPTRVYSTGSGNVFQSPRR ARSVASQSIAYTMSLGAENSVAYSNNIAIPTNFTISVTTEILPVSMK TSVDCTMYICGDSTECNLLLQYGSFCTQLNRALTGIAVEQDKNTQ EVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVT LADAGFIKQYGDCLGDIAARDLCAQKFNGLTVLPPLTDEMIAQYTS ALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKL IANQFNSAIGKIQDLSSTASALGKLQDVVNQNAQALNTLVKQLSSN FGAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEI RASANLAATKMSECVLQGSKRVDFCGKGYHLMSPQSAPHGVVFL HVTYVPAQEKNFHTTAPAICHGDKAHFPREGVFSNGTHWFVTQRNF YEPQIITDNTFVSGNCDVIGIVNNTVYDPLQPELDSFKEELDKYFK NHTSPDVLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGK YEQ	196
spike glycoprotein with a mutation 682-RRAR-685 → 682-QAAQ-685 in the S1-S2	MFVFLVLLPLVSSQCVNLTTTRTQLPPAYTNSFTRGVYYPDKVFRSSV LHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFA STEKSNIIRGWIFGTTLDSTQSLIVNNATNVVIVKCEFQFCNDPFL GVYHKNKSWMESEFRVYSSANNCTFEYVSQPFMDLEGKQGN FKNLREFVFNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINI	191

cleavage site	TRFQTLALHRSYLTGPDSSSGWTAGAAAYVGYLQPRTFLLKYNE NGTITDAVDCALDPLSETKCTLSFTVEKGIYQTSNFRVQPTESIVRF PNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFST FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFRDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLS FELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLP FQQFGRDIADTTDAVTRAGCLIGAEHVNNSEYCDIPGAGICASYQT QTNRPQTLEILDITPCSFSGGVSVITPGTNTSNQVAVLYQDVNCTEV PVAIHADQLTPTWRVYSTGSNVFQSPQQAQSVASQSIIAYTMSLGAE NSVAYSNNISAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECNS LLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFG GFNFSQLPDPSPKSKRSFIEDLLFNKVTADAGFIKQYGDCLGDIAA RDLICAKFNGLTVLPLLTDEMAQYTSALLAGTITSGWTFGAGAAL QIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDLSSTA SALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRDLKVEA EVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQS KRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICH DGKAHFPREGVFSNGTHWFVTQRNFYEPQIITDNTFVSGNCDVV IGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDISGINASVV NIQKEIDRLNEVAKNLNESLIDLQELGKYEYIKWPWYIWLGFIAGLIA IVMVTIMLCCMTSCCCLKGCCSCGSCCKFDEDDSEPVKGVKLHY T	
spike glycoprotein with two proline substitutions (K986P, V987P)	MFVFLVLLPLVSSQCVNLTRTQLPPAYTNSFTRGVYYPDKVFRSSV LHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFA STEKSNIIRGWIFGTTLDSTQSLIVNNATNVVIVKCEFCNDPFL GVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGN FKNLREFVFKNIDGYFKIYSKHTPINLVRDLPPQGSALPLVDLPIGINI TRFQTLALHRSYLTGPDSSSGWTAGAAAYVGYLQPRTFLLKYNE NGTITDAVDCALDPLSETKCTLSFTVEKGIYQTSNFRVQPTESIVRF PNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFST FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFRDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLS FELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLP FQQFGRDIADTTDAVRDPQTLEILDITPCSFSGGVSVITPGTNTSNQVA VLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEH VNNSYECDIPGAGICASYQTQTNPPRRARSVASQSIIAYTMSLGAE NSVAYSNNISAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECNS LLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFG GFNFSQLPDPSPKSKRSFIEDLLFNKVTADAGFIKQYGDCLGDIAA RDLICAKFNGLTVLPLLTDEMAQYTSALLAGTITSGWTFGAGAAL QIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDLSSTA SALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRDLDPPEA EVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQS KRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICH DGKAHFPREGVFSNGTHWFVTQRNFYEPQIITDNTFVSGNCDVV IGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDISGINASVV NIQKEIDRLNEVAKNLNESLIDLQELGKYEYIKWPWYIWLGFIAGLIA IVMVTIMLCCMTSCCCLKGCCSCGSCCKFDEDDSEPVKGVKLHY T	192
spike glycoprotein with four proline substitutions (F817P, A892P, A899P, A942P)	MFVFLVLLPLVSSQCVNLTRTQLPPAYTNSFTRGVYYPDKVFRSSV LHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFA STEKSNIIRGWIFGTTLDSTQSLIVNNATNVVIVKCEFCNDPFL GVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGN	193

	FKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINI TRFQTLLALHRSYLT PGDSSSGW TAGAAAYVGYLQPRTFLLKYNE NGTITDAVDCALDPLSETKCTLSFTVEKGIYQTSNFRVQPTESIVRF PNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFST FKCYGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLS FELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLP FQQFGRDIADTTDAVTRAGCLIGA EHVNNSECDIPIGAGICASYQT QTNRPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEV PVAIHADQLTPTWRVYSTGSNVFQSPRRARSVASQSIIAYTMSLGAE NSVAYSNNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECNS LLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFG GFNFSQILPDPSKPSKRSPIEDLLFNKVT LADAGFIKQYGDCLGDIAA RDLICAKFNGLTVLPPLLTDEMAQYTSALLAGTITSGWTFGAGPAL QIPPMQMAYRFNGIGVTQNVLYENQKLIANQFN SAIGKIQDSLST PSALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRDKVE AEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQ SKRVDFCGKGYHLMSPQSA PHGVVFLHVTVYVPAQEKNTTAPAIC HDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCD VVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVLDGDISGINAS VVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIA LIAIVMTIMLCCMTSCCCLKGCCSCGSCCKFDEDDSEPV LKGVK LHYT	
spike glycoprotein with six proline substitutions (F817P, A892P, A899P, A942P, K986P, V987P)	MFVFLVLLPLVSSQCVNLTRTQLPPAYTNSFTRGVYYPDKVFRSSV LHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFA STEKSNIRGWIFGTTLD SKTQSL LIVN NATNVVIVKCEFCNDPFL GVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGN FKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINI TRFQTLLALHRSYLT PGDSSSGW TAGAAAYVGYLQPRTFLLKYNE NGTITDAVDCALDPLSETKCTLSFTVEKGIYQTSNFRVQPTESIVRF PNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFST FKCYGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLS FELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLP FQQFGRDIADTTDAVTRAGCLIGA EHVNNSECDIPIGAGICASYQT QTNRPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEV PVAIHADQLTPTWRVYSTGSNVFQSPPRRARSVASQSIIAYTMSLGAE NSVAYSNNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECNS LLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFG GFNFSQILPDPSKPSKRSPIEDLLFNKVT LADAGFIKQYGDCLGDIAA RDLICAKFNGLTVLPPLLTDEMAQYTSALLAGTITSGWTFGAGPAL QIPPMQMAYRFNGIGVTQNVLYENQKLIANQFN SAIGKIQDSLST PSALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRDPPE AEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQ SKRVDFCGKGYHLMSPQSA PHGVVFLHVTVYVPAQEKNTTAPAIC HDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCD VVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVLDGDISGINAS VVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIA LIAIVMTIMLCCMTSCCCLKGCCSCGSCCKFDEDDSEPV LKGVK LHYT	194

spike glycoprotein with six proline substitutions (F817P, A892P, A899P, A942P, K986P, V987P) and a 682-RRAR-685 → 682-QQAQ-685 mutation	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVVYPDKVFRSSV LHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFA STEKSNIIRGWIFGTTLDSKTQSLIVNNATNVVIVKCEFCNDPFL GVYYHKNNKSWMESEFRVYSSANNCTFEYVSQLPFLMDLEGKQGN FKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINI TRFQTLLALHRSYLT PGDSSSGWTAGAAAYVGYLQPRTFLLKYNE NGTITDAVDCALDPLSETKCTLSFTVEKGIYQTSNFRVQPTESIVRF PNITNLCPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFST FKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGNNYLYRLFRKSNLKPFRDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLS FELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLP FQQFGRDIADTTDAVTRAGCLIGAEHVNNSECDIPGAGICASYQT QTNRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEV PVAIHADQLTPTWRVYSTGSNVFQSP QQAQ SVASQSIAYTMSLGAE NSVAYSNNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECN LLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFG GFNFSQILPDPSKPSKRS P IEDLLFNKVTADAGFIKQYGDCLGIDIAA RDLICAKFNGLTVLPPLTDEMAQYTSALLAGTITSGWTFGAG P AL QIP P MQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLST P SALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRD P PE AEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQ SKRVDFCGKGYHLSFPQSAPHGVVFLHVTYVPAQEKNFITAPAI HDGKAHFPREGVFSNGTHWVFTQRNFYEPQIITDNTFVSGNCD VVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDIGINAS VVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIA LIAIVMTIMLCCMTSCCCLKGCCSCGSCCKFDEDDSEPVKGVK LHYT	195
Wild type native leader sequence	MFVFLVLLPLVSS	188

[00240] As previously discussed, each of the large sequences are separated by a linker. In some embodiments, the linker is the same linker. In some embodiments, one or more linkers are different. For example, in some embodiments, a different linker is used between each large sequence. As previously discussed, non-limiting examples of linkers include T2A, E2A, P2A, or the like.

[00241] As previously discussed, in certain embodiments, the vaccine delivery system comprises an adenovirus such as but not limited to Ad5, Ad26, Ad35, etc., as well as carriers such as lipid nanoparticles, polymers, peptides, etc.

CD8+ Epitopes

[00242] Examples of methods for identifying potential CD8+ T cell epitopes and screening conservancy of potential CD8+ T cell epitopes are described herein. The present invention is not limited to the particular software systems disclosed, and other software systems are accessible to one of ordinary skill in the art for such methods. The present invention is not limited to the specific haplotypes used herein. For example, one of ordinary skill in the art may select alternative molecules (e.g., HLA molecules) for

molecular docking studies.

[00243] FIG. 10 shows sequence homology analysis for screening conservancy of potential CD8⁺ T cell epitopes, e.g., the comparison of sequence homology for the potential CD8⁺ T cell epitopes among 81,963 SARS-CoV-2 strains (that currently circulate in 190 countries on 6 continents), the 4 major “common cold” Coronaviruses that caused previous outbreaks (e.g., hCoV-OC43, hCoV-229E, hCoV-HKU1-Genotype B, and hCoV-NL63), and the SL-CoVs that were isolated from bats, civet cats, pangolins and camels. Epitope sequences highlighted in yellow present a high degree of homology among the currently circulating 81,963 SARS-CoV-2 strains and at least a 50% conservancy among two or more humans SARS-CoV strains from previous outbreaks, and the SL-CoV strains isolated from bats, civet cats, pangolins and camels.

[00244] From the analysis, 27 CD8⁺ T cell epitopes were selected as being highly conserved. FIG. 11A and FIG. 11B show the docking of the conserved epitopes to the groove of HLA-A*02:01 molecules as well as the interaction scores determined by protein-peptide molecular docking analysis.

[00245] FIG. 12A, FIG. 12B, and FIG. 12C shows that CD8⁺ T cells specific to several highly conserved SARS-CoV-2 epitopes disclosed herein were detected in COVID-19 patients and unexposed healthy individuals. FIG. 13A, FIG. 13B, FIG. 13C, and FIG. 13D shows immunogenicity of the identified SARS-CoV-2 CD8⁺ T cell epitopes.

[00246] The CD8⁺ T cell target epitopes discussed above include S₂₋₁₀, S₁₂₂₀₋₁₂₂₈, S₁₀₀₀₋₁₀₀₈, S₉₅₈₋₉₆₆, E₂₀₋₂₈, ORF1ab₁₆₇₅₋₁₆₈₃, ORF1ab₂₃₆₃₋₂₃₇₁, ORF1ab₃₀₁₃₋₃₀₂₁, ORF1ab₃₁₈₃₋₃₁₉₁, ORF1ab₅₄₇₀₋₅₄₇₈, ORF1ab₆₇₄₉₋₆₇₅₇, ORF7b₂₆₋₃₄, ORF8a₇₃₋₈₁, ORF10₃₋₁₁, and ORF10₅₋₁₃. FIG. 14 shows the genome-wide location of the epitopes. Thus, in certain embodiments, the vaccine composition may comprise one or more CD8⁺ T cell epitopes selected from: S₂₋₁₀, S₁₂₂₀₋₁₂₂₈, S₁₀₀₀₋₁₀₀₈, S₉₅₈₋₉₆₆, E₂₀₋₂₈, ORF1ab₁₆₇₅₋₁₆₈₃, ORF1ab₂₃₆₃₋₂₃₇₁, ORF1ab₃₀₁₃₋₃₀₂₁, ORF1ab₃₁₈₃₋₃₁₉₁, ORF1ab₅₄₇₀₋₅₄₇₈, ORF1ab₆₇₄₉₋₆₇₅₇, ORF7b₂₆₋₃₄, ORF8a₇₃₋₈₁, ORF10₃₋₁₁, ORF10₅₋₁₃, or a combination thereof. Table 3 below describes the sequences for the aforementioned epitope regions.

Table 3

CD8 ⁺ T Cell Epitope	Epitope Sequence	SEQ ID NO:
ORF1ab ₈₄₋₉₂	VMVELVAEL	2
ORF1ab ₁₆₇₅₋₁₆₈₃	YLATALTL	3
ORF1ab ₂₂₁₀₋₂₂₁₈	CLEASFNYL	4
ORF1ab ₂₃₆₃₋₂₃₇₁	WLMWLIINL	5
ORF1ab ₃₀₁₃₋₃₀₂₁	SLPGVFCGV	6

ORF1ab ₃₁₈₃₋₃₁₉₁	FLLNKEMYL	7
ORF1ab ₃₇₃₂₋₃₇₄₀	SMWALIISV	8
ORF1ab ₄₂₈₃₋₄₂₉₁	YLASGGQPI	9
ORF1ab ₅₄₇₀₋₅₄₇₈	KLSYGIATV	10
ORF1ab ₆₄₁₉₋₆₄₂₇	YLDAYNMMI	11
ORF1ab ₆₇₄₉₋₆₇₅₇	LLLDDFVEI	12
S ₂₋₁₀	FVFLVLLPL	13
S ₆₉₁₋₆₉₉	SIIAYTMSL	14
S ₉₅₈₋₉₆₆	ALNTLVKQL	15
S ₉₇₆₋₉₈₄	VLNDILSRL	16
S ₁₀₀₀₋₁₀₀₈	RLQSLQTYV	17
S ₁₂₂₀₋₁₂₂₈	FIAGLIAIV	18
E ₂₀₋₂₈	FLAFVVFL	19
E ₂₆₋₃₄	FLLVTLAIL	20
E ₂₆₋₃₄	FLLNKEMYL	21
M ₅₂₋₆₀	IFLWLLWPV	22
M ₈₉₋₉₇	GLMWLSYFI	23
ORF6 ₃₋₁₁	HLVDFQVTI	24
ORF7b ₂₆₋₃₄	IIFWFSLEL	25

ORF8a ₃₁₋₃₉	YVDDPCPI	26
ORF8a ₇₃₋₈₁	YIDIGNYTV	27
ORF10 ₃₋₁₁	YINVFAFPF	28
ORF10 ₅₋₁₃	NVFAFPFTI	29

[00247] The present invention is not limited to the aforementioned CD8⁺ T cell epitopes. For example, the present invention also includes variants of the aforementioned CD8⁺ T cell epitopes, for example sequences wherein the aforementioned CD8⁺ T cell epitopes are truncated by one amino acid (examples shown below in Table 4).

Table 4

CD8 ⁺ T Cell Epitope Origin:	Sequence with Single AA Truncation	SEQ ID NO:
ORF1ab ₈₄₋₉₂	VMVELVAE	30
ORF1ab ₁₆₇₅₋₁₆₈₃	LATALLTL	31
ORF1ab ₂₂₁₀₋₂₂₁₈	CLEASFNY	32
ORF1ab ₂₃₆₃₋₂₃₇₁	LMWLIINL	33
ORF1ab ₃₀₁₃₋₃₀₂₁	SLPGVFCG	34
ORF1ab ₃₁₈₃₋₃₁₉₁	LLNKEMYL	35
ORF1ab ₃₇₃₂₋₃₇₄₀	SMWALIIS	36
ORF1ab ₄₂₈₃₋₄₂₉₁	LASGGQPI	37
ORF1ab ₅₄₇₀₋₅₄₇₈	KLSYGIAT	38
ORF1ab ₆₄₁₉₋₆₄₂₇	LDAYNMMI	39
ORF1ab ₆₇₄₉₋₆₇₅₇	LLLDDFVE	40

S ₂₋₁₀	VFLVLLPL	41
S ₆₅₁₋₆₅₉	SIIAYTMS	42
S ₉₅₈₋₉₆₆	LNTLVKQL	43
S ₉₇₆₋₉₈₄	VLNDILSR	44
S ₁₀₀₀₋₁₀₀₈	LQSLQTYV	45
S ₁₂₂₀₋₁₂₂₈	FIAGLIAI	46
E ₂₀₋₂₈	LAFVVFL	47
E ₂₆₋₃₄	FLLVTLAI	48
E ₂₆₋₃₄	LLNKEMYL	49
M ₅₂₋₆₀	IFLWLLWP	50
M ₈₉₋₉₇	LMWLSYFI	51
ORF6 ₃₋₁₁	HLVDFQVT	52
ORF7b ₂₆₋₃₄	IFWFSLEL	53
ORF8a ₃₁₋₃₉	YVDDPCP	54
ORF8a ₇₃₋₈₁	IDIGNYTV	55
ORF10 ₃₋₁₁	YINVFAFP	56
ORF10 ₅₋₁₃	VFAFPFTI	57

[00248] The present invention is not limited to the aforementioned CD8⁺ T cell epitopes.

[00249] In certain embodiments, the vaccine composition comprises 1-10 CD8⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 2-10 CD8⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 2-15 CD8⁺ T cell target epitopes. In certain

embodiments, the vaccine composition comprises 2-20 CD8⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 2-30 CD8⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 2-15 CD8⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 2-5 CD8⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 5-10 CD8⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 5-15 CD8⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 5-20 CD8⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 5-25 CD8⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 5-30 CD8⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 10-20 CD8⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 10-30 CD8⁺ T cell target epitopes.

CD4+ Epitopes

[00250] Examples of methods for identifying potential CD4⁺ T cell epitopes and screening conservancy of potential CD4⁺ T cell epitopes are described herein. The present invention is not limited to the particular software systems disclosed, and other software systems are accessible to one of ordinary skill in the art for such methods. The present invention is not limited to the specific haplotypes used herein. For example, one of ordinary skill in the art may select alternative molecules (e.g., HLA molecules) for molecular docking studies.

[00251] FIG. 15 shows the identification of highly conserved potential SARS-CoV-2-derived human CD4⁺ T cell epitopes that bind with high affinity to HLA-DR molecules. Out of a total of 9,594 potential HLA-DR-restricted CD4⁺ T cell epitopes from the whole genome sequence of SARS-CoV-2-Wuhan-Hu-1 strain (MN908947.3), 16 epitopes that bind with high affinity to HLA-DRB1 molecules were selected. The conservancy of the 16 CD4⁺ T cell epitopes was analyzed among human and animal Coronaviruses. Shown are the comparison of sequence homology for the 16 CD4⁺ T cell epitopes among 81,963 SARS-CoV-2 strains (that currently circulate in 6 continents), the 4 major "common cold" Coronaviruses that caused previous outbreaks (i.e. hCoV-OC43, hCoV-229E, hCoV-HKU1, and hCoV-NL63), and the SL-CoVs that were isolated from bats, civet cats, pangolins and camels. Epitope sequences highlighted in green present high degree of homology among the currently circulating 81,963 SARS-CoV-2 strains and at least a 50% conservancy among two or more humans SARS-CoV strains from previous outbreaks, and the SL-CoV strains isolated from bats, civet cats, pangolins and camels.

[00252] From the analysis, 16 CD4⁺ T cell epitopes were selected as being highly conserved. FIG. 16A and FIG. 16B show the docking of the conserved epitopes to the groove of HLA-A*02:01 molecules as well as the interaction scores determined by protein-peptide molecular docking analysis.

[00253] FIG. 17A, FIG. 17B, and FIG. 17C show that CD4⁺ T cells specific to several highly conserved SARS-CoV-2 epitopes disclosed herein were detected in COVID-19 patients and unexposed healthy individuals. FIG. 18A, FIG. 18B, FIG. 18C, and FIG. 18D show immunogenicity of the identified SARS-CoV-2 CD4⁺ T cell epitopes.

[00254] The CD4⁺ T cell target epitopes discussed above include ORF1a₁₃₅₀₋₁₃₆₅, ORF1ab₅₀₁₉₋₅₀₃₃, ORF6₁₂₋₂₆, ORF1ab₆₀₈₈₋₆₁₀₂, ORF1ab₆₄₂₀₋₆₄₃₄, ORF1a₁₈₀₁₋₁₈₁₅, S₁₋₁₃, E₂₆₋₄₀, E₂₀₋₃₄, M₁₇₆₋₁₉₀, N₃₈₈₋₄₀₃, ORF7a₃₋₁₇, ORF7a₁₋₁₅, ORF7b₈₋₂₂, ORF7a₉₈₋₁₁₂, and ORF8₁₋₁₅. FIG. 14 shows the genome-wide location of the epitopes. Thus, in certain embodiments, the vaccine composition may comprise one or more CD4⁺ T cell target epitopes selected from ORF1a₁₃₅₀₋₁₃₆₅, ORF1ab₅₀₁₉₋₅₀₃₃, ORF6₁₂₋₂₆, ORF1ab₆₀₈₈₋₆₁₀₂, ORF1ab₆₄₂₀₋₆₄₃₄, ORF1a₁₈₀₁₋₁₈₁₅, S₁₋₁₃, E₂₆₋₄₀, E₂₀₋₃₄, M₁₇₆₋₁₉₀, N₃₈₈₋₄₀₃, ORF7a₃₋₁₇, ORF7a₁₋₁₅, ORF7b₈₋₂₂, ORF7a₉₈₋₁₁₂, ORF8₁₋₁₅, or a combination thereof. Table 5 below describes the sequences for the aforementioned epitope regions.

Table 5

CD4 ⁺ T Cell Epitope	Epitope Sequence	SEQ ID NO:
ORF1a ₁₃₅₀₋₁₃₆₅	KSAFYILPSIISNEK	58
ORF1a ₁₈₀₁₋₁₈₁₅	ESPFVMMSSAPPAQYE	59
ORF1ab ₅₀₁₉₋₅₀₃₃	PNMLRIMASLVLARK	60
ORF1ab ₆₀₈₈₋₆₁₀₂	RIKVQMLSDTLKNL	61
ORF1ab ₆₄₂₀₋₆₄₃₄	LDAYNMMISAGFSLW	62
S ₁₋₁₃	MFVFLVLLPLVSS	63
E ₂₀₋₃₄	FLAFVVFLVTLAIL	64
E ₂₆₋₄₀	FLLVTLAILTALRLC	65
M ₁₇₆₋₁₉₀	LSYYKLGASQRVAGD	66
ORF6 ₁₂₋₂₆	AEILLIIMRTFKVSI	67
ORF7a ₁₋₁₅	MKIILFLALITLATC	68
ORF7a ₃₋₁₇	IIFLALITLATCEL	69
ORF7a ₉₈₋₁₁₂	SPIFLIVAAIVFITL	70
ORF7b ₈₋₂₂	DFYLCFLAFLFLVL	71

ORF8b ₁₋₁₅	MKFLVFLGIITVAA	72
N ₃₈₈₋₄₀₃₁	KQQTVTLTPAADLDDF	73

[00255] The present invention is not limited to the aforementioned CD4⁺ T cell epitopes. For example, the present invention also includes variants of the aforementioned CD4⁺ T cell epitopes, for example sequences wherein the aforementioned CD4⁺ T cell epitopes are truncated by one or more amino acids or extended by one or more amino acids (examples shown below in Table 6).

Table 6

CD4 ⁺ T Cell Epitope Origin	Sequence with Single AA Truncation	SEQ ID NO:
ORF1a ₁₃₅₀₋₁₃₆₅	KSAFYILPSIISNE	74
ORF1a ₁₈₀₁₋₁₈₁₅	ESPFVMSAPPAQY	75
ORF1ab ₅₀₁₉₋₅₀₃₃	PNMLRIMASLVLR	76
ORF1ab ₆₀₈₆₋₆₁₀₂	RIKVQMLSDLKN	77
ORF1ab ₆₄₂₀₋₆₄₃₄	LDAYNMMISAGFSL	78
S ₁₋₁₃	MFVFLVLLPLVS	79
E ₂₀₋₃₄	FLAFVWFLVTLAI	80
E ₂₆₋₄₀	FLLVTLAILTALRL	81
M ₁₇₆₋₁₉₀	LSYYKLGASQRVAG	82
ORF6 ₁₂₋₂₆	AEILLIMRTFKVS	83
ORF7a ₁₋₁₅	MKIILFLALITLAT	84
ORF7a ₃₋₁₇	IIFLALITLATCE	85

ORF7a ₉₈₋₁₁₂	SPIFLIVAAIVFIT	86
ORF7b ₈₋₂₂	DFYLCFLAFLLLFLV	87
ORF8b ₁₋₁₅	MKFLVFLGIITTVA	88
N ₃₈₈₋₄₀₃₁	KQQTVTLTPAADLDD	89
ORF1a ₁₃₅₀₋₁₃₆₅	SAFYILPSIISNEK	90
ORF1a ₁₈₀₁₋₁₈₁₅	SPFVMMSAPPAQYE	91
ORF1ab ₅₀₁₉₋₅₀₃₃	NMLRIMASLVLARK	92
ORF1ab ₆₀₈₈₋₆₁₀₂	IKVQMLSDTLKNL	93
ORF1ab ₆₄₂₀₋₆₄₃₄	DAYNMMISAGFSLW	94
S ₁₋₁₃	FVFLVLLPLVSS	95
E ₂₀₋₃₄	LAFVVFLVTLAIL	96
E ₂₆₋₄₀	LLVTLAILTALRLC	97
M ₁₇₆₋₁₉₀	SYKLGASQRVAGD	98
ORF6 ₁₂₋₂₆	EILLIIMRTFKVSI	99
ORF7a ₁₋₁₅	KIILFLALITLATC	100
ORF7a ₃₋₁₇	IFLALITLATCEL	101
ORF7a ₉₈₋₁₁₂	PIFLIVAAIVFITL	102
ORF7b ₈₋₂₂	FYLCFLAFLLLFLVL	103
ORF8b ₁₋₁₅	KFLVFLGIITTVAA	104

N ₃₈₈₋₄₀₃₁	QQTVTLPPAADLDDF	105
-----------------------	-----------------	-----

[00256] The present invention is not limited to the aforementioned CD4⁺ T cell target epitopes.

[00257] In certain embodiments, the vaccine composition comprises 1-10 CD4⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 2-10 CD4⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 2-15 CD4⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 2-20 CD4⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 2-30 CD4⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 2-15 CD4⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 2-5 CD4⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 5-10 CD4⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 5-15 CD4⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 5-20 CD4⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 5-25 CD4⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 5-30 CD4⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 10-20 CD4⁺ T cell target epitopes. In certain embodiments, the vaccine composition comprises 10-30 CD4⁺ T cell target epitopes.

B cell Epitopes

[00258] Examples of methods for identifying potential B cell epitopes and screening conservancy of potential B cell epitopes are described herein. The present invention is not limited to the particular software systems disclosed, and other software systems are accessible to one of ordinary skill in the art for such methods.

[00259] FIG. 19 shows the conservation of Spike-derived B cell epitopes among human, bat, civet cat, pangolin, and camel coronavirus strains. Multiple sequence alignment performed using ClustalW among 29 strains of SARS coronavirus (SARS-CoV) obtained from human, bat, civet, pangolin, and camel. This includes 7 human SARS/MERS-CoV strains (SARS-CoV-2-Wuhan (MN908947.3), SARS-HCoV-Urbani (AY278741.1), CoV-HKU1-Genotype-B (AY884001), CoV-OC43 (KF923903), CoV-NL63 (NC005831), CoV-229E (KY983587), MERS (NC019843)); 8 bat SARS-CoV strains (BAT-SL-CoV-WIV16 (KT444582), BAT-SL-CoV-WIV1 (KF367457.1), BAT-SL-CoV-YNLF31C (KP886808.1), BAT-SARS-CoV-RS672 (FJ588686.1), BAT-CoV-RATG13 (MN996532.1), BAT-CoV-YN01 (EPIISL412976), BAT-CoV-YN02 (EPIISL412977), BAT-CoV-19-ZXC21 (MG772934.1)); 3 Civet SARS-CoV strains (SARS-CoV-Civet007 (AY572034.1), SARS-CoV-A022 (AY686863.1), SARS-CoV-B039 (AY686864.1)); 9 pangolin SARS-CoV strains (PCoV-GX-P2V(MT072864.1), PCoV-GX-P5E(MT040336.1), PCoV-GX-P5L (MT040335.1), PCoV-GX-P1E (MT040334.1), PCoV-GX-P4L (MT040333.1), PCoV-MP789 (MT084071.1), PCoV-GX-P3B (MT072865.1), PCoV-Guangdong-P2S (EPIISL410544), PCoV-Guangdong (EPIISL410721)); 4 camel SARS-CoV strains (Camel-CoV-HKU23 (KT368891.1), DcCoV-HKU23 (MN514967.1), MERS-CoV-Jeddah (KF917527.1), Riyadh/RV141 (NC028752.1)) and 1 recombinant

strain (FJ211859.1)). Regions highlighted with blue color represent the sequence homology. The B cell epitopes, which showed at least 50% conservancy among two or more strains of the SARS Coronavirus or possess receptor-binding domain (RBD) specific amino acids were selected as candidate epitopes.

[00260] From the analysis, 22 B cell epitopes were selected as being highly conserved. FIG. 20A and FIG. 20B shows the docking of the conserved epitopes to the ACE2 receptor as well as the interaction scores determined by protein-peptide molecular docking analysis. FIG. 21A, FIG. 21B, FIG. 21C, FIG. 21D, FIG. 21E, FIG. 21F, and FIG. 21G shows immunogenicity of the identified SARS-CoV-2 B cell epitopes.

[00261] The B cell target epitopes discussed above include S₂₈₇₋₃₁₇, S₅₂₄₋₅₉₈, S₆₀₁₋₆₄₀, S₈₀₂₋₈₁₉, S₈₈₈₋₉₀₉, S₃₆₉₋₃₉₃, S₄₄₀₋₅₀₁, S₁₁₃₃₋₁₁₇₂, S₃₂₉₋₃₆₃, S₅₉₋₈₁, and S₁₃₋₃₇. FIG. 2B shows the genome-wide location of the epitopes. Thus, in certain embodiments, the vaccine composition may comprise one or more B cell target epitopes selected from: S₂₈₇₋₃₁₇, S₅₂₄₋₅₉₈, S₆₀₁₋₆₄₀, S₈₀₂₋₈₁₉, S₈₈₈₋₉₀₉, S₃₆₉₋₃₉₃, S₄₄₀₋₅₀₁, S₁₁₃₃₋₁₁₇₂, S₃₂₉₋₃₆₃, S₅₉₋₈₁, and S₁₃₋₃₇. In some embodiments, the B cell epitope is whole spike protein. In some embodiments, the B cell epitope is a portion of the spike protein. Table 7 below describes the sequences for the aforementioned epitope regions.

Table 7

B Cell Epitope	Epitope Sequence	SEQ ID NO:
S ₁₃₋₃₇	SQCVNLTTRTQLPPAYNSFTRGVY	106
S ₅₉₋₈₁	FSNVTWFHAIHVSGTNGTKRFDN	107
S ₂₈₇₋₃₁₇	DAVDCALDPLSETKCTLKSTVEKGIYQTSN	108
S ₆₀₁₋₆₄₀	GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGS	109
S ₅₂₄₋₅₉₈	VCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDI ADTTDAVRDPQTLEILDITPCSFGGVSVI	110

S ₄₄₀₋₅₀₁	NLDSKVGGNYYLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGF NCYFPLQSYGFQPT	111
S ₃₆₉₋₃₉₃	YNSASFSTFKCYGVSP TKLNDLCFT	112
S ₃₂₉₋₃₆₃	FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVA	113
S ₁₁₃₃₋₁₁₇₂	VNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGI	114
S ₈₀₂₋₈₁₉	FSQILPDPSKPSKRSFIE	115
S ₉₈₉₋₉₀₉	FGAGAALQIPFAMQMAYRFNGI	116

[00262] The present invention is not limited to the aforementioned B cell epitopes. For example, the present invention also includes variants of the aforementioned B cell epitopes, for example sequences wherein the aforementioned B cell epitopes are truncated by one or more amino acids or extended by one or more amino acids (examples shown below in Table 8).

Table 8

Origin of Epitope	Sequence with AA Truncation	SEQ ID NO:
S ₁₃₋₃₅	SQCVNLTTRTQLPPAYTNSFTRG	117
S ₅₉₋₇₉	FSNVTWFHAIHVSGTNGTKRF	118
S ₂₈₇₋₃₁₅	DAVDCALDPLSETKCTLSFTVEKGIYQT	119

S ₆₀₁₋₆₃₈	GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYST	120
S ₅₂₄₋₅₉₆	VCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDI ADTTDAVRDPQTLEILDITPCSFGGVS	121
S ₄₄₀₋₄₉₉	NLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGF NCYFPLQSYGFQP	122
S ₃₆₉₋₃₉₁	YNSASFSTFKCYGVSPTKLNDLC	123
S ₃₂₉₋₃₆₁	FPNITNLCPFGEVFNATRFASVYAWNRKRISNC	124
S ₁₁₃₃₋₁₁₇₀	VNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDIS	125
S ₈₀₂₋₈₁₇	FSQILPDPSKPSKRSF	126
S ₈₈₈₋₉₀₇	FGAGAALQIPFAMQMAYRFN	127
S ₁₅₋₃₇	CVNLTTRTQLPPAYTNSFTRGVY	128
S ₆₁₋₈₁	NVTWFHAIHVSGTNGTKRFDN	129
S ₂₈₉₋₃₁₇	VDCALDPLSETKCTLKSFTVEKGIYQTSN	130
S ₆₀₃₋₆₄₀	NTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGS	131

S ₅₂₆₋₅₉₈	GPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIAD TTDAVRDPQTLEILDITPCSFEGGVSVI	132
S ₄₄₂₋₅₀₁	DSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFN CYFPLQSYGFQPTE	133
S ₃₇₁₋₃₉₃	SASFSTFKCYGVSPTKLNDLCFT	134
S ₃₃₁₋₃₆₃	NITNLCPFGEVFNATRFASVYAWNRKRISNCVA	135
S ₁₁₃₅₋₁₁₇₂	NTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGI	136
S ₈₀₄₋₈₁₉	QILPDPSKPSKRSFIE	137
S ₈₅₀₋₉₀₉	AGAALQIPFAMQMAYRFNGI	138

[00263] As previously discussed, in some embodiments, the B cell epitope is in the form of whole spike protein. In some embodiments, the B cell epitope is in the form of a portion of spike protein. In some embodiments, the transmembrane anchor of the spike protein has an intact S1–S2 cleavage site. In some embodiments, the spike protein is in its stabilized conformation. In some embodiments, the spike protein is stabilized with proline substitutions at amino acid positions 986 and 987 at the top of the central helix in the S2 subunit. In some embodiments, the composition comprises a trimerized SARS-CoV-2 receptor–binding domain (RBD). In some embodiments, the trimerized SARS-CoV-2 receptor–binding domain (RBD) sequence is modified by the addition of a T4 fibrin–derived foldon trimerization domain. In some embodiments, the addition of a T4 fibrin–derived foldon trimerization domain increases immunogenicity by multivalent display. FIG. 22 shows a non-limiting example of a spike protein comprising one or more mutations.

[00264] In some embodiments, the spike protein comprises Tyr-489 and Asn-487 (e.g., Tyr-489 and Asn-487 help with interaction with Tyr 83 and Gln-24 on ACE-2). In some embodiments, the spike protein comprises Gln-493 (e.g., Gln-493 helps with interaction with Glu-35 and Lys-31 on ACE-2). In some

embodiments, the spike protein comprises Tyr-505 (e.g., Tyr-505 helps with interaction with Glu-37 and Arg-393 on ACE-2). In some embodiments, the composition comprises a mutation 682-RRAR-685 → 682-QQAQ-685 in the S1-S2 cleavage site.

[00265] In some embodiments, the composition comprises at least one proline substitution. In some embodiments, the composition comprises at least two proline substitutions. For example, the proline substitution may be at position K986 and V987.

[00266] In certain embodiments, the vaccine composition comprises 1-10 B cell target epitopes. In certain embodiments, the vaccine composition comprises 2-10 B cell target epitopes. In certain embodiments, the vaccine composition comprises 2-15 B cell target epitopes. In certain embodiments, the vaccine composition comprises 2-20 B cell target epitopes. In certain embodiments, the vaccine composition comprises 2-30 B cell target epitopes. In certain embodiments, the vaccine composition comprises 2-15 B cell target epitopes. In certain embodiments, the vaccine composition comprises 2-5 B cell target epitopes. In certain embodiments, the vaccine composition comprises 5-10 B cell target epitopes. In certain embodiments, the vaccine composition comprises 5-15 B cell target epitopes. In certain embodiments, the vaccine composition comprises 5-20 B cell target epitopes. In certain embodiments, the vaccine composition comprises 5-25 B cell target epitopes. In certain embodiments, the vaccine composition comprises 5-30 B cell target epitopes. In certain embodiments, the vaccine composition comprises 10-20 B cell target epitopes. In certain embodiments, the vaccine composition comprises 10-30 B cell target epitopes.

[00267] For certain embodiments, the epitopes that are selected may be those that achieve a particular score in a binding assay (for binding to an HLA molecule, for example.) For example, in some embodiments, the epitopes selected have an IC₅₀ score of 250 or less in an ELISA binding assay (e.g., an ELISA binding assay specific for HLA-DR/peptide combination, HLA-A*0201/peptide combination, etc.), or the equivalent of the IC₅₀ score of 250 or less in a different binding assay. Binding assays are well known to one of ordinary skill in the art.

Large Sequence(s) Arrangements

[00268] The large sequences of the compositions described may be arranged in various configurations (see FIG. 23). In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by an ORF1a/b protein or a portion thereof followed by Nucleoprotein or a portion thereof. In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by an ORF1a/b protein or a portion thereof followed by Nucleoprotein or a portion thereof is followed by a membrane (M) or a portion thereof.

[00269] In some embodiments, the large sequences may be arranged such that an ORF1a/b protein or a portion thereof followed by a nucleoprotein (N) or a portion thereof. In some embodiments, the large sequences may be arranged such that an ORF1a/b protein or a portion thereof followed by nucleoprotein

(N) or a portion thereof is followed by a membrane (M) or a portion thereof.

[00270] In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by fragment 1 or a portion thereof. In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by fragment 2 or a portion thereof. In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by fragment 4 or a portion thereof. In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by fragment 5 or a portion thereof. In further embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by fragment 1 or a portion thereof, followed by fragment 5 or a portion thereof.

[00271] In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by a nucleocapsid protein or a portion thereof. In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by a ORF1ab protein or portion thereof, followed by a ORF3 protein or portion thereof followed by an Envelope protein or protein thereof, followed by Membrane protein or portion thereof followed by an ORF6 protein or portion thereof, followed by a ORF7a protein or portion thereof. In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by a membrane protein or portion thereof, followed by a envelope protein or portion thereof, followed by a Nsp3 protein or portion thereof, followed by a Nsp5 protein or portion thereof, followed by a Nsp12 protein or portion thereof.

[00272] In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by one large sequence. In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by two large sequences. In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by three large sequences. In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by four large sequences. In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by five large sequences.

[00273] In some embodiments, the large sequences may be arranged such that a spike glycoprotein (S) or a portion thereof (e.g., the RBD) is followed by one large sequence both are driven each by a promoter or both are driven by a single promoter but separated by a linker as illustrated in FIG x, y and z)

Vaccine Candidates

[00274] As previously discussed, the present invention provides vaccine compositions comprising an antigen featuring: one or more large sequence, two or more large sequences, three or more large

sequences, four or more large sequences, or five or more large sequences. In some embodiment, the large sequences comprise at least one B cell epitope and at least one CD4+ T cell epitope, at least one B cell epitope and at least one CD8+ T cell epitope, at least one CD4+ T cell epitope and at least one CD8+ T cell epitope, or at least one B cell epitope, at least one CD4+ T cell epitope, and at least one CD8+ T cell epitope.

[00275] Table 9 and FIG. 24 shows examples of vaccine compositions described herein. The present invention is not limited to the examples in Table 9.

Table 9:

Vaccine Candidate	Sequence:	SEQ ID NO:
1 <i>promoter, 5'UTR and leader sequence, Spike glycoprotein (HexaPro-mutations), linker, nucleocapsid, Stop Codon, 3'UTR and PolyA tail</i>	<u>CTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT</u> <u>AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGG</u> <u>CCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTTGACGTCAATAAT</u> <u>GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAAT</u> <u>GGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA</u> <u>TCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC</u> <u>GCCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCTACTTTGGCA</u> <u>GTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCA</u> <u>CGTTCGTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCCAATTTT</u> <u>GTATTTATTTATTTTAAATTTTGTGCAGCGATGGGGGCGGGGGGGG</u> <u>GGGGGGGGGCGCGCGCCAGGCGGGGGCGGGGGCGGGGGCGAGGGGCGG</u> <u>GGCGGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCG</u> <u>CGCTCCGAAAGTTTCCCTTTTATGGCGAGGCGGCGGCGGCGGCGGCC</u> <u>TATAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCGCTGCCT</u> <u>TGCGCCCGTGCCCGGCTCCGCGCGCGCCCTCGCGCCGCGCCGCGCCGG</u> <u>CTCTGACTGACCGCGTTACTCCACAGGTGAGCGGGCGGGGACGGGCC</u> <u>TTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAAATGACGGCTTGTTCTT</u> <u>TTCTGTGGCTGCGTGAAAGCCTTGAGGGGGCTCCGGGAGGGGCCCTTG</u> <u>TGCGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTTGTGTGCGTGG</u> <u>GGAGCGCCGCGTGCGGCTCCGCGCTGCCCGGCGGCTGTGAGCGCTG</u> <u>CGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCGAGTGTGCGCGAGGGG</u> <u>AGCGCGGCGGGGGGCGGTGCCCGCGGTGCGGGGGGGGCTGCGAG</u> <u>GGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAG</u> <u>GGGGTGTGGGCGCGTCCGTCGGGCTGCAACCCCCCTGCACCCCCCT</u> <u>CCCCGAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTAC</u> <u>GGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCA</u> <u>GGTGGGGGTGCCGGGCGGGGCGGGGCGGCTCGGGCGGGGAGGG</u> <u>CTCGGGGGAGGGGCGCGGGGCCCCCGGAGCGCGGGCGGGCTGTCC</u> <u>AGGCGCGGCGAGCCGAGCCATTGCCCTTTATGGTAATCGTGCGAGAG</u> <u>GGCGCAGGGACTTCCCTTTGTCCCAATCTGTGCGGAGCGGAAATCTGG</u> <u>GAGGCGCGCGCCGACCCCCCTAGCGGGCGCGGGGCGAAGCGGTGC</u> <u>GGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGC</u> <u>CGCGCCGCGTCCCCCTTCTCCCTCTCCAGCTCGGGGCTGTCCGCGG</u> <u>GGGGACGGCTGCCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCGGCT</u> <u>TCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAACCATGTTTCATG</u> <u>CCTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT</u> <u>GTCTCATCATTTTGGCAAAGAATTGGAGAATAAATAGTATTCTTCGTGC</u> <u>CCACAGACTCAGAGAGAACCCGCCACCATGTTTCGTGTTCTTCTTCTT</u> <u>GCTGCCCCGTGGTGAGCAGCCAGTGCGTGAACCTGACCACGAGACC</u> <u>AGCTGCCCCCGCCTACACCAACAGCTTCACCAGGGGCGTGTACTACC</u>	139

CCGACAAGGTGTTTCAGGAGCAGCGTGCTGCACAGCACCCAGGACCTG TTCCTGCCCTTCTTCAGCAACGTGACCTGGTTCCACGCCATCCACGTGA GCGGCACCAACGGCACCAAGAGGTTTCGACAACCCCGTGCTGCCCTTC AACGACGGCGTGTACTTCGCCAGCACCGAGAAGAGCAACATCATCAGG GGCTGGATCTTCGGCACCAACCCTGGACAGCAAGACCCAGAGCCTGCT GATCGTGAACAACGCCACCAACGTGGTGATCAAGGTGTGCGAGTTCCA GTTCTGCAACGACCCCTTCCTGGGCGTGTAACCACAAGAACAACAA GAGCTGGATGGAGAGCGAGTTTCAGGGTGACAGCAGCGCCAACAACT GCACCTTCGAGTACGTGAGCCAGCCCTTCCTGATGGACCTGGAGGGCA AGCAGGGCAACTTCAAGAACCTGAGGGAGTTCGTGTTCAAGAACATCG ACGGCTACTTCAAGATCTACAGCAAGCACACCCCATCAACCTGGTGAG GGACCTGCCCCAGGGCTTCAGCGCCCTGGAGCCCTGGTGACCTGCG CCATCGGCATCAACATCACCAGGTTCCAGACCCTGCTGGCCCTGCACA GGAGCTACCTGACCCCGGCGACAGCAGCAGCGGCTGGACCGCCGG CGCCGCGCCTACTACGTGGGCTACCTGCAGCCCAGGACCTTCCTGCT GAAGTACAACGAGAACGGCACCATCACCGACGCCGTGGAAGTGCGCCCT GGACCCCTGAGCGAGACCAAGTGACCCCTGAAGAGCTTACCCGTGG AGAAGGGCATCTACCAGACCAGCAACTTCAGGGTGACGCCACCCGAGA GCATCGTGAGGTTCCCAACATCACCAACCTGTGCCCTTCGGCGAGG TGTTCACGCCACCAGGTTCCGCCAGCGTGACGCCTGGAACAGGAAGA GGATCAGCAACTGCGTGGCCGACTACAGCGTGCTGTACAACAGCGCCA GCTTCAGCACCTTCAAGTGCTACGGCGTGAGCCCCACCAAGCTGAACG ACCTGTGCTTACCAACGTGTACGCCGACAGCTTCGTGATCAGGGGCG ACGAGGTGAGGCAGATCGCCCCGGCCAGACCGGCAAGATCGCCGAC TACAACTACAAGCTGCCCGACGACTTCACCGGCTGCGTGATCGCCTGG AACAGCAACAACCTGGACAGCAAGGTGGGCGGCAACTACAACCTACCTG TACAGGCTGTTTCAAGAGAGCAACCTGAAGCCCTTCGAGAGGGACATC AGCACCGAGATCTACCAGGCGCGGACGACCCCTGCAACGGCGTGGA GGGCTTCAACTGCTACTTCCCCCTGCAGAGCTACGGCTTCCAGCCAC CAACGGCGTGGGCTACCAGCCCTACAGGGTGGTGGTGCTGAGCTTCG AGCTGCTGCACGCCCCCGCCACCGTGTGCGGCCCAAGAAGAGCACC AACCTGGTGAAGAACAAGTGCGTGAACCTTCAACTTCAACGGCCTGACC GGCACCGGCGTGCTGACCGAGAGCAACAAGAAGTTCTGCCCTTCCA GCAGTTCGGCAGGGACATCGCCGACACCACCGACGCCGTGAGGGACC CCCAGACCCTGGAGATCCTGGACATCACCCCTGCAGCTTCGGCGGC GTGAGCGTGATACCCCGGACCAACACCAGCAACCAGGTGGCCGT GCTGTACCAGGACGTGAACGTGACCGAGGTGCCGTGGCCATCCACG CCGACAGCTGACCCCACTGGAGGGGTACAGCACCAGGCGAGCAAC GTGTTCCAGACCAGGGCGGCTGCCTGATCGGCGCCGAGCACGTGAA CAACAGCTACGAGTGCGACATCCCCATCGGCGCCGGCATCTGCGCCAG CTACCAGACCCAGACCAACAGCCCCGGCAGCGCGCCAGCGTGGCCA GCCAGAGCATCATCGCTACACCATGAGCCTGGGCGCCGAGAACAGCG TGGCCTACAGCAACAACAGCATCGCCATCCCAACCAACTTCAACATCAG CGTGACCACCGAGATCCTGCCCGTGAGCATGACCAAGACCAGCGTGGA CTGCACCATGTACATCTGCGGCGACAGCACCGAGTGACAGCAACCTGCT GCTGCAGTACGGCAGCTTCTGCACCCAGCTGAACAGGGGCCCTGACCG GCATCGCCGTGGAGCAGGACAAGAACACCCAGGAGGTGTTTCGCCAG GTGAAGCAGATCTACAAGACCCCCCATCAAGGACTTCGGCGGCTTC AACTTCAGCCAGATCCTGCCCGACCCAGCAAGCCAGCAAGAGGAG CCGCATCGAGGACCTGCTGTTCAACAAGGTGACCCCTGGCCGACGCCG GCTTCATCAAGCAGTACGGCGACTGCCTGGGCGACATCGCCGCCAGG GACCTGATCTGCGCCAGAAAGTTCAACGGCCTGACCGTGCTGCCCCCC CTGCTGACCGACGAGATGATCGCCAGTACACCAGCGCCCTGCTGGCC GGCACCATCACCAGCGGCTGGACCTTCGGCGCCGGCGCGCCCTGCA GATCCCCTTCGATGCAGATGGCCTACAGGTTCAACGGCATCGGCGT GACCCAGAAGCTGCTGTACGAGAACCAGAAGCTGATCGCCAACAGTT CAACAGCGCATCGGCAAGATCCAGGACAGCCCTGAGCAGCACCAGCA GCGCCCTGGGCAAGCTGCAGGACGTGGTGAACCAGAACGCCAGGCC	
---	--

	CTGAACACCCTGGTGAAGCAGCTGAGCAGCAACTTCGGCGCCATCAGC AGCGTGCTGAACGACATCCTGAGCAGGCTGGAC<u>CCCCCG</u>GAGGCCGA GGTGAGATCGACAGGCTGATCACCAGCAGGCTGCAGAGCCTGCAGA CCTACGTGACCCAGCAGCTGATCAGGGCCGCCGAGATCAGGGCCAGC GCCAACCTGGCCGCCACCAAGATGAGCGAGTGCGTGCTGGGCCAGAG CAAGAGGGTGACTTCTGCGGCAAGGGCTACCACCTGATGAGCTTCCC CCAGAGCGCCCCCAGCGCTGGTGTTCCTGCACGTGACCTACGTGC CCGCCCAGGAGAAGAACTTCACCACCGCCCCCGCCATCTGCCACGAC GGCAAGGCCCCACTTCCCCAGGGAGGGCGTGTTCTGTGAGCAACGGCAC CCACTGGTTCGTGACCCAGAGGAACCTTCTACGAGCCCCAGATCATCAC CACCGACAACACCTTCGTGAGCGGCAACTGCGACGTGGTGATCGGCAT CGTGAACAACACCGTGACGACCCCCCTGCAGCCCCGAGCTGGACAGCTT CAAGGAGGAGCTGGACAAGTACTTCAAGAACCACACCAGCCCCGACGT GGACCTGGGCGACATCAGCGGCATCAACGCCAGCGTGGTGAACATCCA GAAGGAGATCGACAGGCTGAACGAGGTGGCCAAGAACCTGAACGAGA GCCTGATCGACCTGCAGGAGCTGGGCAAGTACGAGCAGTACATCAAGT GGCCCTGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCATCG TGATGGTGACCATCATGCTGTGCTGCATGACCAGCTGCTGCAGCTGCCT GAAGGGCTGCTGCAGCTGCGGCAGCTGCTGCAAGTTCGACGAGGACG ACAGCGAGCCCGTGCTGAAGGGCGTGAAGCTGCACTACACCGGAAGC <u>GGAGCCACGAACCTCTCTCTGTTAAAGCAAGCAGGAGATGTTGAAGAAA</u> <u>ACCCCGGGCTATGAGCGACAACGGCCCCCAGAACCAGAGGAACGC</u> <u>CCCCAGGATCACCTTCGGCGGCCCCAGCGACAGCACCGGCAGCAAC</u> <u>CAGAACGGCGAGAGGAGCGGCGCCAGGAGCAAGCAGAGGAGGCC</u> <u>CAGGGCCTGCCAACAACACCGCCAGCTGGTTACCGCCCTGACCCA</u> <u>GCACGGCAAGGAGGACCTGAAGTTCCCCAGGGGCCAGGGCGTGCCC</u> <u>ATCAACACCAACAGCAGCCCCGACGACCAGATCGGCTACTACAGGAG</u> <u>GGCCACCAGGAGGATCAGGGGCGGCGACGGCAAGATGAAGGACCTG</u> <u>AGCCCCAGGTGGTACTTCTACTACCTGGGCACCGGCCCGAGGCCGG</u> <u>CCTGCCCTACGGCGCCAACAAGGACGGCATCATCTGGGTGGCCACCG</u> <u>AGGGCGCCCTGAACACCCCCAAGGACCACATCGGCACCAGGAACCC</u> <u>CGCCAACAACGCCGCCATCGTGCTGCAGCTGCCCCAGGGCACCC</u> <u>CTGCCCAAGGGCTTCTACGCCGAGGGCAGCAGGGGCGGCAGCCAGG</u> <u>CCAGCAGCAGGAGCAGCAGCAGGAGCAGGAACAGCAGCAGGAACA</u> <u>GCACCCCCGGCAGCAGCAGGGGCACCCAGCCCCGCCAGGATGGCCGG</u> <u>CAACGGCGGCGACGCCGCCCTGGCCCTGCTGCTGCTGGACAGGCTG</u> <u>AACCAGCTGGAGAGCAAGATGAGCGGCAAGGGCCAGCAGCAGGAGG</u> <u>GCCAGACCGTGACCAAGAAGAGCGCCGCCGAGGCCAGCAAGAAGCC</u> <u>CAGGCAGAAGAGGACCGCCACCAAGGCCTACAACGTGACCCAGGCC</u> <u>TTCCGGCAGGAGGGGCCCGAGCAGACCCAGGGCAACTTCGGCGACC</u> <u>AGGAGCTGATCAGGCAGGGCACCGACTACAAGCACTGGCCCCAGATC</u> <u>GCCCCAGTTGCCCCCAGCGCCAGCGCCTTCTTCGGCATGAGCAGGAT</u> <u>CGGCATGGAGGTGACCCCCAGCGGCACCTGGCTGACCTACACCGGC</u> <u>GCCATCAAGCTGGACGACAAGGACCCCACTTCAAGGACCAGGTGAT</u> <u>CCTGCTGAACAAGCACATCGACGCCTACAAGACCTTCCCCCCCCACCG</u> <u>AGCCCAAGAAGGACAAGAAGAAGAAGGCCGACGAGACCCAGGCCCT</u> <u>GCCCCAGAGGCAGAAGAAGCAGCAGACCGTGACCTGCTGCCCGCC</u> <u>GCCGACCTGGACGACTTCAGCAAGCAGCTGCAGCAGAGCATGAGCA</u> <u>GCGCCGACAGCACCCAGGCC</u><u>TGACTCGAGCTGGTACTGCATGCACGC</u> <u>AATGCTAGCTGCCCCCTTCCCGTCTGGGTACCCCGAGTCTCCCCGA</u> <u>CCTCGGGTCCCAGGTATGCTCCACCTCCACCTGCCCACTCACCAC</u> <u>CTCTGCTAGTTCCAGACACCTCCCAAGCACGCAGCAATGCAGCTCAA</u> <u>AACGCTTAGCCTAGCCACACCCCCACGGGAAACAGCAGTGATTAACC</u> <u>TTTAGCAATAAACGAAAGTTTAACTAAGCTATACTAACCCAGGGTTGG</u> <u>TCAATTCGTGCCAGCCACACCCCTGGAGCTAGCAAAAAAAA</u>	
2	<u>CTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT</u> <u>AGTTCATAGCCCATATATGGAGTTCGCGTTACATAACTTACGGTAAATGG</u>	140

<i>promoter.</i> <i>5'UTR and</i> <i>leader</i> <i>sequence,</i> <i>Spike</i> <i>glycoprotein</i> <i>(HexaPro-</i> <i>mutations),</i> <i>linker.</i> ORF1ab <i>(non-</i> <i>annotated),</i> Stop Codon. 3'UTR and PolyA tail	CCCGCCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAAT GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAAT GGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA TCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC GCCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCCTACTTGGCA GTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCA CGTTCTGCTTCACCTCTCCCCATCTCCCCCCCCCTCCCCACCCCCAATTTT GTATTTATTTATTTTTAATTATTTGTGCAGCGATGGGGGCGGGGGGGG GGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGG GGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCG CGCTCCGAAAGTTTCTTTTATGGCGAGGCGGCGGGCGGCGGGCCCC TATAAAAAGCGAAGCGCGCGGGCGGGGCGGGAGTCGCTGCGCGCTGCCCT TCGCCCCGTGCCCGCTCCGCGCGCGCTCGGCGCGCGCGCGCGCGCGG CTCTGACTGACCGCGTTACTCCACAGGTGAGCGGGCGGGACGGCCC TTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTATGACGGCTTGTTCCT TTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGGCCCTTG TGCGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTTGTGTGTCGTGG GGAGCGCGCGTGCCTGCGGCTCCGCGCTGCCCGGCGGCTGTGAGCGCTG CGGGCGCGGGCGCGGGGCTTTGTGCGCTCCGCAAGTGTGCGCGAGGGG AGCGCGCGCGGGGCGGTGCCCGCGGTGCGGGGGGGGCTGCGAG GGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGCGGGGGGTGAGCAG GGGGTGTGGGCGCGTCCGTGCGGCTGCAACCCCCCTGCAACCCCCCT CCCCGAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTAC GGGGCGTGCGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCA GGTGGGGGTGCCGGGCGGGGCGGGGCGGCTCGGGCGGGGAGGG CTCGGGGGAGGGGCGCGGCGGGCCCCCGGAGCGCGCGCGGCTGTGG AGGCGCGGCGAGCGCGAGCCATTGCCTTTTATGGTAATCGTGCGAGAG GGCGCAGGGACTTCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGG GAGGCGCGCGCGCACCCCCCTAGCGGGCGCGGGGCGAAGCGGTGC GGCGCGGCGAGGAAGGAAATGGGCGGGGAGGGCCCTTCGTGCGTCGC CGCGCGCGCGTCCCTTCTCCCTCTCCAGCCTCGGGGCTGTCCGCGG GGGACGGCTGCCCTCGGGGGGGACGGGCGAGGGCGGGGTTCGGCT TCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAACCATTGTTCTATG CCTTCTTCTTTTCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCT GTCTCATCATTTTGGCAAAGAATTGGAGAATAAACTAGTATTCTTCTGGTC CCCACAGACTCAGAGAGAACCCGCCACCATGTTCTGTGTTCTGGTGCT GCTGCCCCGTGGTGAGCAGCCAGTGCGTGAACCTGACCACCAGGACCC AGCTGCCCCCGCTACACCAACAGCTTCAACAGGGGCGGTACTACC CCGACAAGGTGTTCAAGGAGCAGCGTGTGCACAGCACCCAGGACCTG TTCCTGCCCTTCTTACGAACGTGACCTGGTTCCACGCCATCCACGTGA CGGCGACCAACGGCACCAAGAGGTTGACAACCCCGTGTGCCCTTC AACGACGGCGTGTACTTCGCCAGCACCGAGAAGAGCAACATCATCAGG GGCTGGATCTTCGGCACCAACCTGGACAGCAAGACCCAGAGCCTGCT GATCGTGAACAACGCCACCAACGTGGTGATCAAGGTGTGCGAGTTCCA GTTCTGCAACGACCCCTTCTTGGCGGTGTACTACCACAAGAACAACA GAGCTGGATGGAGAGCGAGTTACGGGTGTACAGCAGCGCCAACAAC GCACCTTCGAGTACGTGAGCCAGCCCTTCTGATGGACCTGGAGGGCA AGCAGGGCAACTTCAAGAACCTGAGGGAGTTCGTGTTCAAGAACATCG ACGGCTACTTCAAGATCTACAGCAAGCACACCCCCATCAACCTGGTGAG GGACCTGCCCCAGGGCTTCAGCGCCCTGGAGCCCCCTGGTGACCTGC CCATCGGCATCAACATCACCAGGTTCCAGACCCTGCTGGCCCTGCACA GGAGCTACCTGACCCCCGGCGACAGCAGCAGCGGCTGGACCGCCGG CGCCGCGCCTACTACGTGGGCTACCTGCAGCCCAGGACCTTCTGCT GAAGTACAACGAGAACGGCACCATCACCGACGCCGTGGACTGCGCCCT GGACCCCTGAGCGAGACCAAGTGACCCCTGAAGAGCTTACCCGTGG AGAAGGGCATCTACCAGACCAACTTCAGGGTGACGCCACCGAGA GCATCGTGAGGTTCCCAACATCACCAACCTGTGCCCTTCGGCGAGG TGTTCAACGCCACAGGTTGCCAGCGTGACGCTGGAACAGGAAGA
---	--

	GGATCAGCAACTGCGTGGCCGACTACAGCGTGCTGTACAACAGCGCCA GCTTCAGCACCTTCAAGTGCTACGGCGTGAGCCCCACCAAGCTGAACG ACCTGTGCTTCACCAACGTGTACGCCGACAGCTTCGTGATCAGGGGCG ACGAGGTGAGGCAGATCGCCCCCGGCCAGACCGGCAAGATCGCCGAC TACAACTACAAGCTGCCGACGACTTCACCGGCTGCGTGATCGCCTGG AACAGCAACAACCTGGACAGCAAGGTGGGCGGCAACTACAACCTACCTG TACAGGCTGTTTACAGGAAGAGCAACCTGAAGCCCTTCGAGAGGGACATC AGCACCGAGATCTACCAGGCCGCGCAGCACCCCTGCAACGGCGTGGA GGGCTTCAACTGCTACTTCCCCCTGCAGAGCTACGGCTTCAGCCCCAC CAACGGCGTGGGCTACCAGCCCTACAGGGTGGTGGTGCTGAGCTTCG AGCTGCTGCACGCCCCCGCCACCGTGTGCGGCCCAAGAAGAGCACC AACCTGGTGAAGAACAAGTGCGTGAACCTTCAACTTCAACGGCCCTGACC GGCACCGGCGTGCTGACCGAGAGCAACAAGAAGTTCCTGCCCTTCCA GCAGTTCGGCAGGGACATCGCCGACACCACCGACGCCGTGAGGGACC CCCAGACCCTGGAGATCCTGGACATCACCCCTGCAGCTTCGGCGGC GTGAGCGTGATCACCCCGGCCACCAACACCAGCAACCAGGTGGCCGT GCTGTACCAGGACGTGAACCTGCACCGAGGTGCCCGTGGCCATCCACG CCGACCAGCTGACCCCCACCTGGAGGGTGTACAGCACCGGCAGCAAC GTGTTCCAGACCAGGGCCGGCTGCCTGATCGGCGCCGAGCACGTGAA CAACAGCTACGAGTGCGACATCCCCATCGGCGCCGGCATCTGCGCCAG CTACCAGACCCAGACCAACAGCCCCGGCAGCGCCAGCAGCGTGGCCA GCCAGAGCATCATCGCTACACCATGAGCCTGGGCGCCGAGAACAGCG TGGCCTACAGCAACAACAGCATCGCCATCCCCACCAACTTCACCATCAG CGTGACCACCGAGATCCTGCCCGTGAGCATGACCAAGACCAGCGTGGA CTGCACCATGTACATCTGCGGCGACAGCACCGAGTGCAGCAACCTGCT GCTGCAGTACGGCAGCTTCTGCACCCAGCTGAACAGGGCCCTGACCG GCATCGCCGTGGAGCAGGACAAGAACACCCAGGAGGTGTTGCCCCAG GTGAAGCAGATCTACAAGACCCCCCCCCATCAAGGACTTCGGCGGCTTC AACTTCAGCCAGATCCTGCCCGACCCAGCAAGCCAGCAAGAGGAG CCCCATCGAGGACCTGCTGTTCAACAAGGTGACCCTGGCCGACGCCG GCTTCATCAAGCAGTACGGCGACTGCCTGGGCGACATCGCCGCCAGG GACCTGATCTGCGCCAGAAAGTTCAACGGCCCTGACCGTGCTGCCCCCC CTGCTGACCGACGAGATGATCGCCAGTACACCAGCGCCCTGCTGGCC GGCACCATCACAGCGGCTGGACCTTCGGCGCCGGCCCGGCCCTGCA GATCCCCCTTCGCGATGACAGATGGCCTACAGGTTCAACGGCATCGGCCG GACCCAGAACGTGCTGTACGAGAACCAGAAAGCTGATCGCCAACCAAGT CAACAGCGCCATCGGCAAGATCCAGGACAGCCTGAGCAGCACCCCGCA GCGCCCTGGGCAAGCTGCAGGACGTGGTGAACCAGAACGCCCCAGGCC CTGAACACCCCTGGTGAAGCAGCTGAGCAGCAACTTCGGCGCCATCAGC AGCGTGCTGAACGACATCCTGAGCAGGCTGGACCCCCGAGGCCGA GGTGAGATCGACAGGCTGATCACCGGCAGGCTGCAGAGCCTGCAGA CCTACGTGACCCAGCAGCTGATCAGGGCCGCGAGATCAGGGCCAGC GCCAACCTGGCCGCCACCAAGATGAGCGAGTGCCTGCTGGGCCAGAG CAAGAGGGTGGACTTCTGCGGCAAGGGCTACCACCTGATGAGCTTCCC CCAGAGCGCCCCCACGGCGTGCTGTTTCTGACAGTGACCTACGTGC CCGCCCAGGAGAAGAACTTACCACCGCCCCCGCCATCTGCCACGAC GGCAAGGCCCCACTTCCCCAGGGAGGGCGTGTTCTGTGAGCAACGGCAC CCACTGGTTCTGTACCCAGAGGAACCTTCTACGAGCCCCAGATCATCAC CACCGACAACACCTTCGTGAGCGGCAACTGCGACGTGGTGATCGGCAT CGTGAACAACACCGTGTACGACCCCTGCAGCCCGAGCTGGACAGCTT CAAGGAGGAGCTGGACAAGTACTTCAAGAACCACACCAGCCCCGACGT GGACCTGGGCGACATCAGCGGCATCAACGCCAGCGTGGTGAACATCCA GAAGGAGATCGACAGGCTGAACGAGGTGGCCAAGAACCTGAACGAGA GCCTGATCGACCTGCAGGAGCTGGGCAAGTACGAGCAGTACATCAAGT GGCCCTGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCATCG TGATGGTGACCATCATGCTGTGCTGCATGACCAGCTGCTGCAGTGCCT GAAGGGCTGCTGCAGCTGCGGCAGCTGCTGCAAGTTCGACGAGGACG ACAGCGAGCCCGTGCTGAAGGGCGTGAAGCTGCACTACACCGGAAGC	
--	---	--

	GGAGCCACGAACCTTCTCTCTGTTAAAGCAAGCAGGAGATGTTGAAGAAA ACCCCGGGGCTCAAACCACTGAAACAGCWCACCTCTTGTAAATGTTAAC CGCTTTAATGTGGCTATTACAAGAGCAAAAATTGGCATTGTTGTGCATAA TGTCTGACAGAGATCTTTATGACAAGCTGCAATTCACAAGTCTAGAAG TACCGCGTCGTAACGTGGCTACATTACAAGCGGAAAATGTAAGTGGAC TCTTTAAGGACTGTAGTAAGATCATAACTGGTCTTCATCCTACACAAGC ACCTACACACCTTAGTGTTGATACAAAATTCAAGACTGAGGGACTATGT GTTGACATACCAGGCATWCCWAAGGACATGACCTATMGWAGACTCAT CTCYATGATGGGTTTCAAATGAATTAYCAAGTTAATGGTTACCTAAYA TGTTYATCACCCGYGARGAAGCCATMMGMCAYGWCGTGCATGGATT GGCTTTGATGTAGAGGGKTGTCATGCTACTAGGGATGCTGTCCGGTACT AACCTACCTCTCCAGTTAGGATTTTCTACAGGTGTTAACTTAGTAGCTG TACCAACTGGCTATGTTGACACTGAAAACAATACAGAATTCACCAGAG TTAATGCAAAACCTCCACCAGGTGACCAATTTAAACATCTTATACCAC TATGTACAAAGGTTTACCCTGGAACATAGTGCCTATCAAGATAGTACAA ATGCTCAGTGATACACTGAAAGGATTATCRGACAGAGTTGTGTTGTCC TATGGGCACATGGCTTTGAACTTACATCAATGAAGTACTTTGTCAAGAT TGGACCTGAAAGAACGTGTTGTCTGTGTGACAAACGTGCAACTGTTT TTCTACTTCATCAGACAATTATGCCTGCTGGAACATTCTGTGGGTTT GACTATGTCTATAATCCATTATGATTGATGTCCAGCAGTGGGGTTTTAC AGGTAACCTTCAGAGTAATCACGATCAGCATTGCCAAGTGCATGGCAA CGCTCATGTGGCTAGTTGTGATGCTATCATGACTAGATGTTTAGCAGTC CATGAGTGCTTTGTTAAGCGCGTTGACTGGTCTGTTGAGTACCCAATTA TAGGTGATGAAGTGAAGATCAATGCCGCATGCAGAAAAGTGCAACATA TGGTTGTAAAGTCTGCATTGCTTGCTGACAAATCCCAGTTCTTCATGA CATTGGAACCCAAAGGCTATCAAATGTGTCCRCAGGCTGAAGTGG ATTGGAAGTTCTATGATGCTCAGCCCTGCAGTGACAAAGCTTATAAAAT AAAAGAAGTCTTCTATTCTTATGCTACACATCATGATAAATTCATTGATG GTGTTTGTATTTTGAATTTGTAACGTTGATCGTTACCCTGCCAATGCT ATTGTRTGCAGGTTGACACGAGAGTCTTGCTCAAATTTGAAGTGGCA GGTTGTGATGGTGGTAGTTGTATGTAAATAAGCATGCATTCCACACTC CAGCTTTTGATAAAAGTGCATTTACTAATTTAAAGCAATTGCCTTTCTTT TATTACTCTGACAGTCCCTGTGAGTCACATGGCAAGCAGGTTGTTCTG ACATTGATTATGTACCACTCAAATCTGCTACRTGTATAACACGATGCAAT TTGGGRGGTGCTGTTTGCAGACATCATGCAAATGAGTACCGACAGTAC TTGGATGCATACAATATGATGATTCTGCTGGCTTTAGCCTCTGGATTTA CAACAGTTTGACACTTATAACCTGTGGAACACCTTTACCAGTTACA GAGTTTAGAAAATGTGGCTTACAATGTTGTTAACAAAGGACACTTCGAT GGACAAGCTGGTGAAGCACCTGTTCCGTCATTAATAATGTTGTTTACA CAAAGGTAGATGGTGGTATGATGATGAGATCTTTGAAAACAAGACAACAC TTCCTGTTAATGTTGCATTTGAGCTTTGGGCTAAGCGTAACATTAACC AGTGCCAGAGATTAAGATACTCAATAATTTGGGTGTCGATATCGCTGCT AATACTGTAATCTGGGACTACAAGAGAGAAGCACCAGCACATATGTCA ACAATAGGTGTCTGCACAATGACTGACATTGCCAAGAAACCTACTGAG AGTGCTTGTTCCCTCGCTTACTGTCTTATTTGATGGTAGAGTGGAAGGAC AGGTAGACCTTTTAGAAATGCCCGTAATGGTGTGTTAATAACAGAAGG TTCAGTTAAAGGTTTAATACCTTCAAAGGGACCAGCACAAGCTAGTGT CAATGGAGTCACATTAATTGGAGAATCAGTAAAAACACAGTTTAATTAT TTAAGAAAGTAGATGGCATCATTCAACAGTTGCCTGAAACCTACTTTA CTCAGAGCCGAGACTTAGAGGATTTCAAGCCCAGATCACAAATGGAA ACTGACTTTCTTGAGCTCGCTATGGATGAATCATACAACGGTACAAGC TTGAAGGCTATGCCCTCGAACATATCGTTTATGGAGATTTTAGTCATGG ACAGCTTGGTGGACTTCATCTAATGATTGGTCTAGCTAAGCGCTCACA AGATTACCACTTAAATTAGAGGATTTATCCCTACGGACAGTACAGTG AAAAATTATTTATAACAGATGCGCAAACAGGTTTCATCAAAATGCGTGT GCTCTGTTATTGATCTTCTGCTTATGACTTTGTTGAGATAATAAAGTCA CAAGATTTATCAGTGGTTTCAAAGGTGGTCAAAGTCACAATTGACTATG CTGAAATTTCAATTCATGTTATGGTGAAGGATGGACATGTTGAAACCTT	
--	--	--

	TTACCCAAAATTACAAGCGAGTCAGGCGTGGCAACCAGGAGTTGCAA TGCCTAACTTGTATAAGATGCAGAGAATGCTTCTTGAAAAATGTGACCT TCAGAATTATGGTGAAAAATGCTGTCAATACCAAARGGAATAATGATGAAT GTCGCAAAATATACTCAACTGTGTCAATATTTAAATACACTYACATTAGC YGTGCCATATAATATGAGAGTTATCCATTTTGGTGCTGGCTCRGACAAA GGAGTTGCACCCGGCACAGCTGTTCTCAGACAGTGGTTGCCAATTGG CACACTACTTGTGATTGATCTTAACGACTTCGTCTCTGACGCTGAT TCCACTCTAATTGGAGACTGTGCAACCGTACATACAGCTAACAAATGG GATCTCATTATTAGCGATATGTATGATCCTAAAACCAAACACGTGACAA AGGAAAATGATTCAAAAGAAGGATTTTCACTTACCTGTGTGGATTAT TAAACAAAAATTAGCCCTGGGAGGCTCTGTGGCTGTAAAGATAACTGA GCATTCTTGGAATGCGGATCTCTACAAGCTCATGGGACATTTCTCATGG TGGACAGCTTTTGTACAAATGTTAATGCATCTTCATCAGAAGCATTTT AATTGGAGTTAACTATCTTGGTAAGCCAAAAGAACAATTTGATTGTTAC ACCATGCATGCTAACTACATTTTCTGGAGGAATACAAACCCGATTCAAT TGTCTTCTATTCACTTTTGTACATGAGTAAGTTCCTCTTAAATTAAGG GGAACAGCTGTCATGCTTTAAAGGAGAACCAAATCAATGAAATGATT TATTCTCTACTTGAAAAAGGCAGACTTATCATTAGGGAAAACAACAGA GTTGTTGTCTCAAGTGATGTTCTTGTTAATACTAAACGAACATGACTC <u>GAGCTGGTACTGCATGCACGCAATGCTAGCTGCCCTTTCCCGTCTG</u> <u>GGTACCCCGAGTCTCCCCGACCTCGGGTCCCAGGTATGCTCCACCC</u> <u>TCCACCTGCCCACTCACCACCTCTGCTAGTTCCAGACACCTCCCAAG</u> <u>CACGCAGCAATGCAGCTCAAAACGCTTAGCCTAGCCACACCCCCACG</u> <u>GGAAACAGCAGTGATTAACCTTTAGCAATAAACGAAAGTTTAACATAAG</u> <u>CTATACTAACCCAGGGTTGGTCAATTTCTGTGCCAGCCACACCCCTGGA</u> <u>GCTAGCAAAAAAAA</u>	
3 <i>promoter,</i> 5'UTR and leader sequence, Spike glycoprotein (HexaPro- mutations), linker, Envelope Membrane, ORF7a, Stop Codon, 3'UTR and PolyA tail	<u>CTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT</u> <u>AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGG</u> <u>CCCGCCTGGCTGACCGCCCAACGACCCCGCCCATTTGACGTCAATAAT</u> <u>GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAAT</u> <u>GGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA</u> <u>TCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC</u> <u>GCCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCTACTTTGGCA</u> <u>GTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCA</u> <u>CGTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCAATTTT</u> <u>GTATTTATTTATTTTTTAATTTATTTGTGCAGCGATGGGGGCGGGGGGG</u> <u>GGGGGGGGCGCGCGCCAGGCGGGGGCGGGGGCGGGGGCGAGGGGGGG</u> <u>GGCGGGGGCGAGGCGGAGAGGTGCGGGCGGCGAGCAATCAGAGCGGGC</u> <u>CGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGGCGGGCGGGCGGGCCC</u> <u>TATAAAAGCGAAGCGCGCGGGCGGGGAGTCCGTGCGCGCTGCCT</u> <u>TCGCCCCGTGCCCCGCTCCGCGCGCCGCTCGCGCCCGCCGCCCCGG</u> <u>CTCTGACTGACCGCGTTACTCCACAGGTGAGCGGGCGGGACGGGCC</u> <u>TTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAAATGACGGCTTGTTCTT</u> <u>TTCTGTGGCTGCGTGAAAGCCTTGAGGGGGCTCCGGGAGGGGCCCTTG</u> <u>TGCGGGGGGAGCGGCTCGGGGGGTGCGTGCCTGTGTGTGTGCGTGG</u> <u>GGAGCGCCGCGTGCGGCTCCGCGCTGCCCGGCGGCTGTGAGCGCTG</u> <u>CGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCAAGTGTGCGCGAGGGG</u> <u>AGCGCGGCGGGGGCGGTGCCCGCGGTGCGGGGGGGGCTGCGAG</u> <u>GGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAG</u> <u>GGGGTGTGGGCGCGTCCGTGCGGGCTGCAACCCCCCTGCACCCCCCT</u> <u>CCCCGAGTTTCTGAGCACGGCCCCGGCTTCGGGTGCGGGGCTCCGTAC</u> <u>GGGGCGTGGCGCGGGGCTCGCCGTGCCGGCGGGGGGTGGCGGCA</u> <u>GGTGGGGGTGCCGGGCGGGGCGGGGCGGGGCTCGGGCGGGGGAGGG</u> <u>CTCGGGGGAGGGGCGCGGGGCCCCCGGAGCGCGCGGGGCTGTGCG</u> <u>AGGCGCGGCGAGCCGCGAGCCATTGCCTTTTATGGTAATCGTGCAGAG</u> <u>GGCGCAGGGACTTCCCTTTGTCCCAATCTGTGCGGAGCGGAAATCTGG</u> <u>GAGGCGCGCGCGACCCCCCTAGCGGGCGCGGGGCGAAGCGGTGC</u>	141

	<u>GGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGC</u> <u>CGCGCCGGCGTCCCCCTTCTCCCTCTCCAGCCTCGGGGCTGTCCGCGG</u> <u>GGGGACGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCCGGCT</u> <u>TCTGGCGTGTGACCGGGCGGCTCTAGAGCCTCTGCTAACCATGTTTCATG</u> <u>CCTTCTTCTTTTTCTTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCT</u> <u>GTCTCATCATTTTGGCAAAGAATTGGAGAATAAACTAGTATTCTTCTGGT</u> <u>CCACAGACTCAGAGAGAACCCGCCACCATGTTCTGTGTTCTGGTGCT</u> <u>GCTGCCCCCTGGTGAGCAGCCAGTGCGTGAACCTGACCACCAGGACCC</u> <u>AGCTGCCCCCGCTACACCAACAGCTTCACCAGGGGCGTGACTACC</u> <u>CCGACAAGGTGTTTCAGGAGCAGCGTGCTGCACAGCACCACAGGACCTG</u> <u>TTCCTGCCCTTCTTCAGCAACGTGACCTGGTTCCACGCCATCCACGTGA</u> <u>GCGGCACCAACGGCACCAAGAGGTTTCGACAACCCCGTGCTGCCCTTC</u> <u>AACGACGGCGTGACTTCGCCAGCACCGAGAAGAGCAACATCATCAGG</u> <u>GGCTGGATCTTCGGCACCAACCCTGGACAGCAAGACCCAGAGCCCTGCT</u> <u>GATCGTGAACAACGCCACCAACGTGGTGATCAAGGTGTGCGAGTTCCA</u> <u>GTTCTGCAACGACCCCTTCTGGGCGTGACTACCACAAGAACAACAA</u> <u>GAGCTGGATGGAGAGCGAGTTTCAGGGTGACAGCAGCGCCAACAACT</u> <u>GCACCTTCGAGTACGTGAGCCAGCCCTTCCTGATGGACCTGGAGGGCA</u> <u>AGCAGGGCAACTTCAAGAACCCTGAGGGAGTTCTGTTCAGAACATCG</u> <u>ACGGCTACTTCAAGATCTACAGCAAGCACACCCCCATCAACCTGGTGAG</u> <u>GGACCTGCCCCAGGGCTTCAGCGCCCTGGAGCCCTGGTGACCTGC</u> <u>CCATCGGCATCAACATCACCAGGTTCCAGACCCTGCTGGCCCTGCACA</u> <u>GGAGCTACCTGACCCCGGCGACAGCAGCAGCGGCTGGACCGCCGG</u> <u>CGCCGCCGCTACTACGTGGGCTACCTGCAGCCCAGGACCTTCCTGCT</u> <u>GAAGTACAACGAGAACGGCACCATCACCGACGCGGTGGACTGCGCCCT</u> <u>GGACCCCTTGAGCGAGACCAAGTGACCCCTGAAGAGCTTCACCGTGG</u> <u>AGAAGGGCATCTACCAGACCAGCAACTTCAGGGTGACGCCACCGAGA</u> <u>GCATCGTGAGGTTCCCAACATCACCAACCTGTGCCCTTCGGCGAGG</u> <u>TGTTCAACGCCACCAGGTTCCGCCAGCGTGACGCTGGAACAGGAAGA</u> <u>GGATCAGCAACTGCGTGGCCGACTACAGCGTGCTGTACAACAGCGCCA</u> <u>GCTTCAGCACCTTCAAGTGCTACGGCGTGAGCCCCACCAAGCTGAACG</u> <u>ACCTGTGCTTCACCAACGTGTACGCCGACAGCTTCGTGATCAGGGGCG</u> <u>ACGAGGTGAGGCAGATCGCCCCCGGCCAGACCGGCAAGATCGCCGAC</u> <u>TACAACTACAAGCTGCCCCGACGACTTCACCGGCTGCGTGATCGCCTGG</u> <u>AACAGCAACAACCTGGACAGCAAGGTGGCGGGCAACTACAACCTACCTG</u> <u>TACAGGCTGTTACGGAAGAGCAACCTGAAGCCCTTCGAGAGGGACATC</u> <u>AGCACCGAGATCTACCAGGCCGGCAGCACCCCTGCAACGGCGTGGA</u> <u>GGGCTTCAACTGCTACTTCCCCCTGCAGAGCTACGGCTTCAGCCAC</u> <u>CAACGGCGTGGGCTACCAGCCCTACAGGGTGGTGGTGCTGAGCTTCG</u> <u>AGCTGCTGCACGCCCCCGCCACCGTGTCGCGCCCCAAGAAGAGCACC</u> <u>AACCTGGTGAAGAACAAGTGCGTGAACCTTCAACTTCAACGGCCTGACC</u> <u>GGCACCGGCGTGCTGACCGAGAGCAACAAGAAGTTCCTGCCCTTCCA</u> <u>GCAGTTCGGCAGGGACATCGCCGACACCACCGACGCGTGAGGGACC</u> <u>CCCAGACCCTGGAGATCCTGGACATACCCCTGCAGCTTCGGCGGC</u> <u>GTGAGCGTGATACCCCGGCACCAACACCAGCAACCAGGTGGCCGT</u> <u>GCTGTACCAGGACGTGAACGCACCGAGGTGCCCGTGCCATCCACG</u> <u>CCGACCAGCTGACCCCACTGGAGGGTGACAGCACCGGCGAGCAAC</u> <u>GTGTTCCAGACCAGGGCCGGCTGCCTGATCGGCGCCGAGCAGTGAA</u> <u>CAACAGCTACGAGTGCGACATCCCCATCGGCGCCGGCATCTGCGCCAG</u> <u>CTACCAGACCCAGACCAACAGCCCCGGCAGCGCCAGCAGCGTGGCCA</u> <u>GCCAGAGCATCATCGCTACACCATGAGCCTGGGCGCCGAGAACAGCG</u> <u>TGGCCTACAGCAACAACAGCATCGCCATCCCACCAACTTCACCATCAG</u> <u>CGTGACCACCGAGATCCTGCCCGTGAGCATGACCAAGACCAGCGTGGA</u> <u>CTGCACCATGTACATCTGCGGCGACAGCACCGAGTGACAGCAACCTGCT</u> <u>GCTGCAGTACGGCAGCTTCTGCACCCAGCTGAACAGGGCCCTGACCG</u> <u>GCATCGCCGTGGAGCAGGACAAGAACACCCAGGAGGTGTTCCGCCAG</u> <u>GTGAAGCAGATCTACAAGACCCCCCATCAAGGACTTCGGCGGCTTC</u> <u>AACTTCAGCCAGATCCTGCCCGACCCAGCAAGCCAGCAAGAGGAG</u>	
--	---	--

<u>CCCCATCGAGGACCTGCTGTTCAACAAGGTGACCCTGGCCGACGCCG</u> <u>GCTTCATCAAGCAGTACGGCGACTGCCTGGGCGACATCGCCGCCAGG</u> <u>GACCTGATCTGCGCCCAGAAGTTCAACGGCCTGACCGTGCTGCCCCC</u> <u>CTGCTGACCGACGAGATGATCGCCAGTACACCAGCGCCCTGCTGGCC</u> <u>GGCACCATCACCAGCGGCTGGACCTTCGGCGCCGGCCCGCCCTGCA</u> <u>GATCCCCTTCCCCATGCAGATGGCCTACAGGTTCAACGGCATCGGCGT</u> <u>GACCCAGAACGTGCTGTACGAGAACCAGAAGCTGATCGCCAACCAAGT</u> <u>CAACAGCGCCATCGGCAAGATCCAGGACAGCCTGAGCAGCACCCCGCA</u> <u>GCGCCCTGGGCAAGCTGCAGGACGTGGTGAACCAGAACGCCAGGCC</u> <u>CTGAACACCCTGGTGAAGCAGCTGAGCAGCAACTTCGGCGCCATCAGC</u> <u>AGCGTGCTGAACGACATCCTGAGCAGGCTGGACCCCCGAGGCCGA</u> <u>GGTGAGATCGACAGGCTGATCACCGGCAGGCTGCAGAGCCTGCAGA</u> <u>CCTACGTGACCCAGCAGCTGATCAGGGCCGCCGAGATCAGGGCCAGC</u> <u>GCCAACCTGGCCGCCACCAAGATGAGCGAGTGCGTGCTGGGCCAGAG</u> <u>CAAGAGGGTGGACTTCTGCGGCAAGGGCTACCACCTGATGAGCTTCCC</u> <u>CCAGAGCGCCCCCAGGGCGTGGTGTTCCTGCACGTGACCTACGTGC</u> <u>CCGCCCAGGAGAAGAACTTACCACCGCCCCCGCCATCTGCCACGAC</u> <u>GGCAAGGCCCACTTCCCCAGGGAGGGCGTGTTCTGAGCAACGGCAC</u> <u>CCACTGGTTCGTGACCCAGAGGAACCTTCTACGAGCCCCAGATCATCA</u> <u>CACCGACAACACCTTCGTGAGCGGCAACTGCGACGTGGTGATCGGCAT</u> <u>CGTGAACAACACCGTGACGACCCCTGCAGCCCGAGCTGGACAGCTT</u> <u>CAAGGAGGAGCTGGACAAGTACTTCAAGAACCACACAGCCCCGACGT</u> <u>GGACCTGGGCGACATCAGCGGCATCAACGCCAGCGTGGTGAACATCCA</u> <u>GAAGGAGATCGACAGGCTGAACGAGGTGGCCAAGAACCTGAACGAGA</u> <u>GCCTGATCGACCTGCAGGAGCTGGGCAAGTACGAGCAGTACATCAAGT</u> <u>GGCCCTGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCATCG</u> <u>TGATGGTGACCATCATGCTGTGCTGCATGACCAGCTGCTGCAGCTGCCT</u> <u>GAAGGGCTGCTGCAGCTGCGGCAGCTGCTGCAAGTTCGACGAGGACG</u> <u>ACAGCGAGCCCGTGCTGAAGGGCGTGAAGCTGCACTACCCGGAAGC</u> <u>GGAGCCACGAACCTTCTCTCTGTTAAAGCAAGCAGGAGATGTTGAAGAA</u> <u>ACCCCGGGCCTATGTACAGCTTCGTGAGCGAGGAGACCGGCACCCTG</u> <u>ATCGTGAACAGCGTGCTGCTGTTCCCTGGCCTTCGTGGTGTTCCTGCTG</u> <u>GTGACCTGGCCATCCTGACCGCCCTGAGGCTGTGCGCCTACTGCTG</u> <u>CAACATCGTGAACGTGAGCCTGGTGAAGCCAGCTTCTACGTGTACA</u> <u>GCAGGGTGAAGAACCTGAACAGCAGCAGGGTGCCCGACCTGCTGGT</u> <u>GGGAAGCGGAGCCACGAACCTTCTCTCTGTTAAAGCAAGCAGGAGATGT</u> <u>TGAAGAAAACCCCGGGCCTATGGCCGACAGCAACGGCACCATCACCGT</u> <u>GGAGGAGCTGAAGAAGCTGCTGGAGCAGTGGAACCTGGTGATCGGCT</u> <u>TCCTGTTCTGACCTGGATCTGCCTGCTGCAGTTCCGCTACGCCAACA</u> <u>GGAACAGGTTCTGTACATCATCAAGCTGATCTTCCTGTGGCTGCTGTG</u> <u>GCCCGTGACCTGGCCTGCTTCGTGCTGGCCGCCGTGTACAGGATCAA</u> <u>CTGGATCACCGGCGGCATCGCCATCGCCATGGCCTGCCCTGGTGGGCT</u> <u>GATGTGGCTGAGCTACTTCATCGCCAGCTTCAGGCTGTTGGCCAGGAC</u> <u>CAGGAGCATGTGGAGCTTCAACCCCGAGACCAACATCCTGCTGAACGT</u> <u>GCCCCCTGCACGGCACCATCCTGACCCAGGCCCTGCTGGAGAGCGAGC</u> <u>TGGTGATCGGCGCCGTGATCCTGAGGGGCCACCTGAGGATCGCCGGC</u> <u>CACCACCTGGGCAGGTGCGACATCAAGGACCTGCCCAAGGAGATCACC</u> <u>GTGGCCACCGAGCAGGACCTGAGCTACTACAAGCTGGGCGCCAGCCA</u> <u>GAGGGTGGCCGGCGACAGCGGCTTCGGCGCCTACAGCAGGTACAGGA</u> <u>TCGGCAACTACAAGCTGAACACCGACACAGCAGCAGCAGCGACAACA</u> <u>TCGCCCTGCTGGTGCAGGGAAGCGGAGCCACGAACCTTCTCTGTTAA</u> <u>AGCAAGCAGGAGATGTTGAAGAAAACCCCGGGCCTATGAAGATCATCC</u> <u>TGTTCTGGCCCTGATCACCTGGCCACCTGCGAGCTGTACCACTACC</u> <u>AGGAGTGCGTGAGGGGCACCACCGTGAGACTCGAGCTGGTACTGCA</u> <u>GCACGCAATGCTAGCTGCCCTTTCCCGTCCCTGGGTACCCCGAGTCTC</u> <u>CCCCGACCTCGGGTCCCAGGTATGCTCCACCTCCACCTGCCCACT</u> <u>CACCACCTCTGCTAGTTCCAGACACCTCCCAAGCACGCAGCAATGCA</u> <u>GCTCAAAACGCTTAGCCTAGCCACACCCCCACGGGAAACAGCAGTGA</u>	
---	--

	<u>TTAACCTTTAGCAATAAACGAAAGTTTAACTAAGCTATACTAACCCAG GGTIGGTCAATTTTCGTGCCAGCCACACCCTGGAGCTAGCAAAAAA</u>	
4 <i>promoter, 5'UTR and leader sequence, Spike glycoprotein (HexaPro- mutations), linker, Envelope, Membrane, Stop Codon, 3'UTR and PolyA tail</i>	<u>CTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT AGTTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGG CCCGCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAAT GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAAT GGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA TCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC GCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCTACTTTGGCA GTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCA CGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCCAATTTT GTATTTATTTATTTTTAATTTATTTGTGCAGCGATGGGGCGGGGGGGGG GGGGGGGGCGCGCCAGGCGGGGGCGGGGGCGGGGGCGAGGGGGCGG GGCGGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGGC CGCTCCGAAAGTTTCTTTTATGGCGAGGGCGGGCGGGCGGGCGGGCCC TATAAAAAGCGAAGCGCGCGGGCGGGCGGGAGTCGCTGCGCGCTGCCCT TCGCCCCGTGCCCCGCTCCGCGCGCGCTCGCGCGCGCGCGCGCGCGG CTCTGACTGACCGCGTTACTCCACAGGTGAGCGGGCGGGACGGCCCC TTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAATGACGGCTTGTTCTT TTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGGCCCTTTG TGCGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTTGTGTGCGTG GGAGCGCGCGGTGCGGGCTCCGCGCTGCCCGGGCGGGGTGTGAGCGCTG CGGGCGCGGGCGGGGGCTTTGTGCGCTCCGCAAGTGTGCGCGAGGGG AGCGCGCGCGGGGGCGGTGCCCCGCGGTGCGGGGGGGGCTGCGAG GGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAG GGGGTGTGGGCGCGTCCGTCCGGCTGCAACCCCCCTGCACCCCCCT CCCCGAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTAC GGGGCGTGGCGCGGGGCTCGCCGTGCCGGCGGGGGGTGGCGGCA GGTGGGGGTGCCGGGCGGGGCGGGGCGGCTCGGGCGGGGAGGG CTCGGGGGAGGGGGCGGGCGGGCCCCCGGAGCGCGGGCGGGCTGTCC AGGCGCGGCGAGCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAG GGCGCAGGGACTTCTTTGTCCCAATCTGTGCGGAGCCGAAATCTGG GAGGCGCGCGCGCACCCCCCTAGCGGGCGGGGGCGAAGCGGTGC GGCGCGGGCAGGAAGGAATGGGCGGGGAGGGCCTTCTGTGCGTCCG CGCGCGCGCTCCCTTCTCCCTCTCCAGCCTCGGGGGCTGTCCGGCG GGGACGGCTGCCCTCGGGGGGGACGGGGCAGGGCGGGGTTCCGGCT TCTGGCGTGTGACCGGGCGGCTCTAGAGCCTCTGCTAACCATGTTTCATG CCTTCTTTTCTTCTACAGCTCTGGGCAACGTGCTGGTTATTGTGCT GTCTCATCATTTTGGCAAAGAATTGAGAATAAACTAGTATTCTTCTGGTC CCACAGACTCAGAGAGAACCCGCCACCATGTTCTGTGTTCTTGGTGCT GCTGCCCCCTGGTGAGCAGCCAGTGCGTGAACCTGACCACAGGACCC AGCTGCCCCCGCCTACACCAACAGCTTCACCAGGGGCGGTGACTACC CCGACAAGGTGTTCAAGAGCAGCGTGCTGCACAGCACCCAGGACCTG TTCCTGCCCTTCTTACGAACGTGACCTGGTTCCACGCCATCCACGTGA GCGGCACCAACGGCACCAAGAGGTTGACAACCCCGTGTGCCCTTC AACGACGGCGTGACTTCGCCAGCACCGAGAAGAGCAACATCATCAGG GGCTGGATCTTCGGCACCCCTGGACAGCAAGACCCAGAGCCTGCT GATCGTGAACAACGCCACCAACGTGGTGATCAAGGTGTGCGAGTTCCA GTTCTGCAACGACCCCTTCTTGGCGGTGACTACCACAAGAACAACAA GAGCTGGATGGAGAGCGAGTTCAGGGGTGTACAGCAGCGCCAACAAC GCACCTTCGAGTACGTGAGCCAGCCCTTCTGATGGACCTGGAGGGCA AGCAGGGCAACTTCAAGAACCTGAGGGAGTTCGTGTTCAAGAACATCG ACGGCTACTTCAAGATCTACAGCAAGCACACCCCCATCAACCTGGTGAG</u>	142

	GGACCTGCCCCAGGGCTTCAGCGCCCTGGAGCCCCTGGTGGACCTGC CCATCGGCATCAACATCACCAGGTTCCAGACCCTGCTGGCCCTGCACA GGAGCTACCTGACCCCCGGCGACAGCAGCAGCGGCTGGACCGCCGG CGCCGCCGCTACTACGTGGGCTACCTGCAGCCCAGGACCTTCCTGCT GAAGTACAACGAGAACGGCACCATCACCGACGCCGTGGACTGCGCCCT GGACCCCCTGAGCGAGACCAAGTGCACCCTGAAGAGCTTCACCGTGG AGAAGGGCATCTACCAGACCAGCAACTTCAGGGTGACGCCACCGAGA GCATCGTGAGGTTCCCCAACATCACCAACCTGTGCCCTTCGGCGAGG TGTTCAACGCCACCAGGTTCCGCCAGCGTGACGCCTGGAACAGGAAGA GGATCAGCAACTGCGTGGCCGACTACAGCGTGCTGTACAACAGCGCCA GCTTCAGCACCTTCAAGTGCTACGGCGTGAGCCCCACCAAGCTGAACG ACCTGTGCTTCACCAACGTGTACGCCGACAGCTTCGTGATCAGGGGCG ACGAGGTGAGGCAGATCGCCCCGGCCAGACCGGCAAGATCGCCGAC TACAACTACAAGCTGCCCCGACGACTTCACCGGCTGCGTGATCGCCTGG AACAGCAACAACCTGGACAGCAAGGTGGGCGGCAACTACAACCTACCTG TACAGGCTGTTACAGGAAGAGCAACCTGAAGCCCTTCGAGAGGGACATC AGCACCGAGATCTACCAGGCCGGCAGCACCCCCTGCAACGGCGTGGA GGGCTTCAACTGCTACTTCCCCCTGCAGAGCTACGGCTTCAGCCAC CAACGGCGTGGGCTACAGCCCTACAGGGTGGTGGTGGTGGTGGTGGT AGCTGCTGCACGCCCCCGCCACCGTGTCGGCCCCAAGAAGAGCACC AACCTGGTGAAGAACAAGTGCGTGAACCTTCAACTTCAACGGCCGTGACC GGCACCGGCGTGCTGACCGAGAGCAACAAGAAGTTCCTGCCCTTCCA GCAGTTCGGCAGGGACATCGCCGACACCACCGACGCCGTGAGGGACC CCCAGACCCTGGAGATCCTGGACATCACCCCCTGCAGCTTCGGCGGC GTGAGCGTGATCACCCCCGGCACCAACACCAGCAACCAGGTGGCCGT GCTGTACCAGGACGTGAACCTGCACCGAGGTGCCCGTGCCATCCACG CCGACCAGCTGACCCCCACCTGGAGGGTGTACAGCACCGGCAGCAAC GTGTTCCAGACCAGGGCCGGCTGCCTGATCGGCGCCGAGCACGTGAA CAACAGCTACGAGTGCGACATCCCCATCGGCGCCGGCATCTGCGCCAG CTACCAGACCCAGACCAACAGCCCCGGCAGCGCGCCAGCGTGGCCA GCCAGAGCATCATCGCCTACACCATGAGCCTGGGCGCCGAGAACAGCG TGGCCTACAGCAACAACAGCATCGCCATCCCCACCAACTTCACCATCAG CGTGACCACCGAGATCCTGCCCGTGAGCATGACCAAGACCAAGCGTGGA CTGCACCATGTACATCTGCGGCGACAGCACCGAGTGACGCAACCTGCT GCTGCAGTACGGCAGCTTCTGCACCCAGCTGAACAGGGCCCTGACCG GCATCGCCGTGGAGCAGGACAAGAACACCCAGGAGGTGTTCCGCCAG GTGAAGCAGATCTACAAGACCCCCCCCCATCAAGGACTTCGGCGGCTTC AACTTCAGCCAGATCCTGCCCGACCCCAGCAAGCCAGCAAGAGGAG CCCCATCGAGGACCTGCTGTTCAACAAGGTGACCCTGGCCGACGCCG GCTTCATCAAGCAGTACGGCGACTGCCTGGGCGACATCGCCGCCAGG GACCTGATCTGCGCCCAAGTTCAACGGCCGTGACCGTGCTGCCCCC CTGCTGACCGACGAGATGATCGCCCAGTACACCAGCGCCCTGCTGGCC GGCACCATCACAGCGGCTGGACCTTCGGCGCCGGCCCGCCCTGCA GATCCCCTTCCCATGCAGATGGCCTACAGGTTCAACGGCATCGGCGT GACCCAGAACGTGCTGTACGAGAACCAGAAGCTGATCGCCAACCAAGT CAACAGCGCCATCGGCAAGATCCAGGACAGCCTGAGCAGCACCCCGCA GCGCCCTGGGCAAGCTGCAGGACGTGGTGAACCAGAACGCCAGGCC CTGAACACCCTGGTGAAGCAGCTGAGCAGCAACTTCGGCGCCATCAGC AGCGTGCTGAACGACATCCTGAGCAGGCTGGACCCCCGAGGCCGA GGTGCAGATCGACAGGCTGATCACCGGCAGGCTGCAGAGCCTGCAGA CCTACGTGACCCAGCAGCTGATCAGGGCCGCCGAGATCAGGGCCAGC GCCAACCTGGCCGCCACCAAGATGAGCGAGTGCGTGCTGGGCCAGAG CAAGAGGGTGGACTTCTGCGGCAAGGGCTACCACCTGATGAGCTTCCC CCAGAGCGCCCCCAGCGCGTGCTGTTCTGCACGTGACCTACGTGC CCGCCCAGGAGAAGAACTTACCACCGCCCCCGCCATCTGCCACGAC GGCAAGGCCCCACTTCCCAGGGAGGGCGTGTTCTGTAGCAACGGCAC CCACTGGTTCTGTACCCAGAGGAATTTCTACGAGCCCCAGATCATCAC CACCGACAACACCTTCGTGAGCGGCAACTGCGACGTGGTGATCGGCAT	
--	---	--

	CGTGAACAACACCGTGACGACCCCTGCAGCCCGAGCTGGACAGCTT CAAGGAGGAGCTGGACAAGTACTTCAAGAACCACACCAGCCCCGACGT GGACCTGGGCGACATCAGCGGCATCAACGCCAGCGTGGTGAACATCCA GAAGGAGATCGACAGGCTGAACGAGGTGGCCAAGAACCCTGAACGAGA GCCTGATCGACCTGCAGGAGCTGGGCAAGTACGAGCAGTACATCAAGT GGCCCTGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCATCG TGATGGTGACCATCATGCTGTGCTGCATGACCAGCTGCTGCAGCTGCCT GAAGGGCTGCTGCAGCTGCGGCAGCTGCTGCAAGTTCGACGAGGACG ACAGCGAGCCCGTGCTGAAGGGCGTGAAGCTGCACTACACCGGAAGC <u>GGAGCCACGAACCTCTCTCTGTTAAAGCAAGCAGGAGATGTTGAAGAAA</u> <u>ACCCCGGGCCTATGGCCGACAGCAACGGCACCATCACCGTGGAGGAG</u> <u>CTGAAGAAGCTGCTGGAGCAGTGGAACTGGTGATCGGCTTCCTGTTT</u> <u>CTGACCTGGATCTGCCTGCTGCAGTTTCGCTACGCCAACAGGAACAGG</u> <u>TTCTGTACATCATCAAGCTGATCTTCCTGTGGCTGCTGTGGCCCGTGA</u> <u>CCCTGGCCTGCTTCGTGCTGGCCGCCGTGTACAGGATCAACTGGATCA</u> <u>CCGGCGGCATCGCCATCGCCATGGCCTGCCTGGTGGGCCTGATGTGG</u> <u>CTGAGCTACTTCATCGCCAGCTTCAGGCTGTTCCGCCAGGACCAGGAGC</u> <u>ATGTGGAGCTTCAACCCGAGACCAACATCCTGCTGAACGTGCCCTG</u> <u>CACGGCACCATCCTGACCAGGCCCTGCTGGAGAGCGAGCTGGTGAT</u> <u>CGGCGCCGTGATCCTGAGGGGCCACCTGAGGATCGCCGGCCACCACC</u> <u>TGGGCAGGTGCGACATCAAGGACCTGCCCAAGGAGATCACCGTGGCC</u> <u>ACCAGCAGGACCCTGAGCTACTACAAGCTGGGCGCCAGCCAGAGGGT</u> <u>GGCCGGCGACAGCGGCTTCGCCGCCCTACAGCAGGTACAGGATCGGCA</u> <u>ACTACAAGCTGAACACCCGACCACAGCAGCAGCAGCGACAACATCGCCC</u> <u>TGCTGGTGCAGGGAAGCGGAGCCACGAACCTCTCTCTGTTAAAGCAAG</u> <u>CAGGAGATGTTGAAGAAAACCCCGGGCCTATGTACAGCTTCGTGAGCG</u> <u>AGGAGACCCGGCACCCCTGATCGTGAACAGCGTGCTGCTGTTCTGGCC</u> <u>TTCTGGTGTTCCTGCTGGTGACCCCTGGCCATCCTGACCGCCCTGAGG</u> <u>CTGTGCGCCTACTGCTGCAACATCGTGAACGTGAGCCTGGTGAAGCC</u> <u>CAGCTTCTACGTGTACAGCAGGGTGAAGAACCCTGAACAGCAGCAGGG</u> <u>TGCCCAGCCTGCTGGTGTGACTCGAGCTGGTACATGACACGCAAT</u> <u>GCTAGCTGCCCCCTTCCCGTCCCTGGGTACCCCGAGTCTCCCCGACCT</u> <u>CGGGTCCCAGGTATGCTCCACCTCCACCTGCCCACTCACCCACTCT</u> <u>GCTAGTTCAGACACCTCCCAAGCACGCAGCAATGCAGCTCAAAACG</u> <u>CTTAGCCTAGCCACACCCCCACGGGAAACAGCAGTGATTAACCTTTAG</u> <u>CAATAAACGAAAGTTTAACTAAGCTATACTAACCCCAAGGGTTGGTCAAT</u> <u>TTCTGCCAGCCACACCTGGAGCTAGCAAAAAAAAAA</u>	
5 <i>promoter.</i> 5'UTR and leader sequence, Spike glycoprotein (HexaPro- mutations), linker. NSP3 NSP5 NSP6 Stop Codon. 3'UTR and PolyA tail	CTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGG CCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAAT GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAAT GGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA TCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC GCCTGGCATTTATGCCAGTACATGACCTTATGGGACTTTCTTACTTGGCA GTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCA CGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCCAATTTT GTATTTATTTATTTTTAATTATTTGTGCAGCGATGGGGCGGGGGGGGG GGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGG GGCGGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGGCG CGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGGCGGGCGGGCC TATAAAAAGCGAAGCGCGCGGCGGGCGGGGAGTCGCTGCGCGCTGCCCT TCGCCCCGTGCCCGCTCCGCGCGCGCTCGCGCGCGCGCGCGCGCGG CTCTGACTGACCGCGTTACTCCCAACAGGTGAGCGGGCGGGACGGCCC TTCTCCTCCGGGCTGTAATTAGCGCTTGGTTAATGACGGCTTGTTCCTT TTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTG TGCGGGGGGAGCGGGCTCGGGGGGTGCGTGCCTGTGTGTGTGCGTGG GGAGCGCCGCGTGCGGCTCCGCGCTGCCCGGCGGCTGTGAGCGCTG	143

	<u>CGGGCGCGGGCGGGGGCTTTGTGCGCTCCGCAGTGTGCGCGAGGGG</u> <u>AGCGCGGCCCGGGGGCGGTGCCCCGCGGTGCGGGGGGGGCTGCCGAG</u> <u>GGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAG</u> <u>GGGGTGTGGGGCGCGTCGGTTCGGGCTGCAACCCCCCTGCACCCCCCT</u> <u>CCCCGAGTTGCTGAGCACGGCCCCGGCTTCGGGTGCGGGGCTCCGTAC</u> <u>GGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCA</u> <u>GGTGGGGGTGCCGGGCGGGGCGGGGCGGCCCTCGGGCGCGGGGAGGG</u> <u>CTCGGGGGAGGGGCGCGGGGCCCCGGAGCGCGCGGGCTGTGCG</u> <u>AGGCGCGGCGAGCCGCGAGCCATTGCCTTTTATGGTAATCGTGCGAGAG</u> <u>GGCGCAGGGACTTCCTTTGTCCCAATCTGTGCGGAGCCGAAATCTGG</u> <u>GAGGCGCCGCGCACCCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGC</u> <u>GGCGCCGCGCAGGAAGGAAATGGGCGGGGAGGGGCTTCGTGCGTCCG</u> <u>CGCGCCGCGCTCCCTTCTCCCTCTCCAGCCTCGGGGCTGTCCGCGG</u> <u>GGGAGCGGCTGCCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCCGGT</u> <u>TCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAACCATGTTTCATG</u> <u>CCTTCTTCTTTTCTACAGCTCCTGGGCAACGTGCTGTTATTGTGCT</u> <u>GTCTCATCATTTTGGCAAAGAATTGAGAATAAACTAGTATTCTTCTGGTC</u> <u>CCCACAGACTCAGAGAGAACCCGCCACCATGTTGTTCTTCTGGTGCT</u> <u>GCTGCCCCCTGGTGAGCAGCCAGTGCCTGAACCTGACCACAGGACCC</u> <u>AGCTGCCCCCGCCTACACCAACAGCTTCACCAGGGGCGTGTAATACC</u> <u>CCGACAAGGTGTTTCAGGAGCAGCGTGCTGCACAGCACCCAGGACCTG</u> <u>TTCCTGCCCTTCTTCAGCAACGTGACCTGGTTCCACGCCATCCACGTGA</u> <u>GCGGCACCAACGGCACCAAGAGTTTCGACAACCCCGTGCTGCCCTTC</u> <u>AACGACGGCGTGACTTCGCCAGCACCGAGAAGAGCAACATCATCAGG</u> <u>GGCTGGATCTTCGGCACCAACCTGGACAGCAAGACCCAGAGCCTGCT</u> <u>GATCGTGAACAACGCCACCAACGTGGTGATCAAGGTGTGCGAGTTCCA</u> <u>GTTCTGCAACGACCCCTTCTGGGCGTGTAATACCACAAGAACAACAA</u> <u>GAGCTGGATGGAGAGCGAGTTTCAGGGGTGACAGCAGCGCCAACT</u> <u>GCACCTTCGAGTACGTGAGCCAGCCCTTCTGATGGACCTGGAGGGCA</u> <u>AGCAGGGCAACTTCAAGAACCTGAGGGAGTTCGTGTTCAAGAACATCG</u> <u>ACGGCTACTTCAAGATCTACAGCAAGCACACCCCATCAACCTGGTGAG</u> <u>GGACCTGCCCCAGGGCTTCAGCGCCCTGGAGCCCCTGGTGACCTGC</u> <u>CCATCGGCATCAACATCACCAAGTTCCAGACCCCTGCTGGCCCTGCACA</u> <u>GGAGCTACCTGACCCCCGGCGACAGCAGCAGCGGCTGGACCGCCGG</u> <u>CGCCGCCGCTACTACGTGGGCTACCTGCAGCCAGGACCTTCCTGCT</u> <u>GAAGTACAACGAGAACGGCACCATCACCGACGCCGTGGACTGCGCCCT</u> <u>GGACCCCTGAGCGAGACCAAGTGACCCCTGAAGAGCTTCACCGTG</u> <u>AGAAGGGCATCTACCAGACCAAGCAACTTCAGGGTGACGCCACCCGAGA</u> <u>GCATCGTGAGGTTCCCCAACATCACCAACCTGTGCCCTTCGGCGAGG</u> <u>TGTTCAACGCCACCAAGTTTCGCCAGCGTGACGCCTGGAACAGGAAGA</u> <u>GGATCAGCAACTGCGTGGCCGACTACAGCGTGCTGTACAACAGCGCCA</u> <u>GCTTCAGCACCTTCAAGTGCTACGGCGTGAGCCCCACCAAGCTGAACG</u> <u>ACCTGTGCTTCACCAACGTGTACGCCGACAGCTTCGTGATCAGGGGCG</u> <u>ACGAGGTGAGGCAGATCGCCCCCGGCCAGACCGGCAAGATCGCCGAC</u> <u>TACAAC TACAAGCTGCCCCGACGACTTCACCGGCTGCGTGATCGCCCTGG</u> <u>AACAGCAACAACCTGGACAGCAAGGTGGGCGGCAACTACAAC TACCTG</u> <u>TACAGGCTGTTACAGGAAGAGCAACCTGAAGCCCTTCGAGAGGGACATC</u> <u>AGCACCGAGATCTACCAGGCCGGCAGCACCCCTGCAACGGCGTGGA</u> <u>GGGCTTCAACTGCTACTTCCCCCTGCAGAGCTACGGCTTCAGCCAC</u> <u>CAACGGCGTGGGCTACCAGCCCTACAGGGTGGTGGTGCTGAGCTTCG</u> <u>AGCTGCTGCACGCCCCCGCCACCGTGTGCGGCCCCAAGAAGAGCACC</u> <u>AACCTGGTGAAGAACAAGTGCGTGAACCTTCAACTTCAACGGCCGTGACC</u> <u>GGCACCGGCGTGCTGACCGAGAGCAACAAGAAGTTCTGCCCTTCCA</u> <u>GCAGTTTCGGCAGGGACATCGCCGACACCCGACGCGGTGAGGGAC</u> <u>CCCAGACCCTGGAGATCCTGGACATCACCCCTGCAGCTTCGGCGGC</u> <u>GTGAGCGTGATCACCCCGGCCACCAACACCAAGCAACAGGTGGCCGT</u> <u>GCTGTACCAGGACGTGAACCTGCACCGAGGTGCCCGTGCCCATCCACG</u> <u>CCGACCAGCTGACCCCCACCTGGAGGGTGTACAGCACCGGCAGCAAC</u>	
--	---	--

GTGTTCCAGACCAGGGCCGGCTGCCTGATCGGGCGCCGAGCACGTGAA CAACAGCTACGAGTGCGACATCCCCATCGGCGCCGGCATCTGCGCCAG CTACCAGACCCAGACCAACAGCCCCGGCAGCGCCAGCAGCGTGGCCA GCCAGAGCATCATCGCCTACACCATGAGCCTGGGCGCCGAGAACAGCG TGGCCTACAGCAACAACAGCATCGCCATCCCCACCAACTTCAACATCAG CGTGACCACCGAGATCCTGCCCCGTGAGCATGACCAAGACCAGCGTGGA CTGCACCATGTACATCTGCGGCGACAGCACCGAGTGACAGCAACCTGCT GCTGCAGTACGGCAGCTTCTGCACCCAGCTGAACAGGGGCCCTGACCG GCATCGCCGTGGAGCAGGACAAGAACACCCAGGAGGTGTTGCCCCAG GTGAAGCAGATCTACAAGACCCCCCATCAAGGACTTCGGCGGCTTC AACTTCAGCCAGATCCTGCCCCAGCCCCAGCAAGCCCAGCAAGAGGAG CCCCATCGAGGACCTGCTGTTCAACAAGGTGACCTTGCCGACGCCG GCTTCATCAAGCAGTACGGCGACTGCCTGGGCGACATCGCCGCCAGG GACCTGATCTGCGCCCAAGATTCAACGGCCTGACCGTGCTGCCCCCC CTGCTGACCGACGAGATGATCGCCCAGTACACCAGCGCCCTGCTGGCC GGCACCATCACCAGCGGCTGGACCTTCGGCGCCGGCCCGGCCCTGCA GATCCCCCTTCATGAGATGGCCTACAGGTTCAACGGCATCGGCGT GACCCAGAACGTGCTGTACGAGAACCAGAAGCTGATCGCCAACCAAGTT CAACAGCGCCATCGGCAAGATCCAGGACAGCCTGAGCAGCACCCCGCA GCGCCCTGGGCAAGCTGCAGGACGTGGTGAACCAGAACGCCAGGCC CTGAACACCCCTGGTGAAGCAGCTGAGCAGCAACTTCGGCGCCATCAGC AGCGTGCTGAACGACATCCTGAGCAGGCTGGACCCCCGAGGCCGA GGTGCAGATCGACAGGCTGATCACCGGCAGGCTGCAGAGCCTGCAGA CCTACGTGACCCAGCAGCTGATCAGGGCCGCCGAGATCAGGGCCAGC GCCAACCTGGCCGCCACCAAGATGAGCGAGTGCGTGCTGGGCCAGAG CAAGAGGGTGGAATCTTGCGGCAAGGGCTACCACCTGATGAGCTTCCC CCAGAGCGCCCCCAGGGCGTGGTGTTCCTGCACGTGACCTACGTGC CCGCCCAGGAGAAGAATTACACACCGCCCCCGCCATCTGCCACGAC GGCAAGGCCCACTTCCCCAGGGAGGGCGTGTTCGTGAGCAACGGCAC CCACTGGTTCTGTGACCCAGAGGAATTCTACGAGCCCCAGATCATCAC CACCGACAACACCTTCTGTGAGCGGCAACTGCGACGTGGTGATCGGCAT CGTGAACAACACCGTGTACGACCCCCCTGCAGCCCCGAGCTGGACAGCTT CAAGGAGGAGCTGGACAAGTACTTCAAGAACCACACACAGCCCCGACGT GGACCTGGGCGACATCAGCGGCATCAACGCCAGCGTGGTGAACATCCA GAAGGAGATCGACAGGCTGAACGAGGTGGCCAAGAACCTGAACGAGA GCCTGATCGACCTGCAGGAGCTGGGCAAGTACGAGCAGTACATCAAGT GGCCCTGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCATCG TGATGGTGACCATCATGCTGTGCTGCATGACCAGCTGCTGCAGCTGCCT GAAGGGCTGCTGCAGCTGCGGCAGCTGCTGCAAGTTCGACGAGGACG ACAGCGAGCCCGTGTGAAGGGCGTGAAGCTGCACTACACCGGAAGC GGAGCCACGAATCTCTCTCTGTTAAAGCAAGCAGGAGATGTTGAAGAAA ACCCGGGGCTGCCCCCACCAGGTGACCTTCGGCGACGACACCGTG ATCGAGGTGCAGGGCTACAAGAGCGTGAAATCACCTTCGAGCTGGAC GAGAGGATCGACAAGGTGCTGAACGAGAAGTGACGCGCCTACACCGT GGAGCTGGGCAACCGAGGTGAACGAGTTTCGCTGCGTGGTGGCCGACG CCGTGATCAAGACCTGCAGCCCGTGAGCGAGCTGCTGACCCCGCTG GGCATCGACCTGGACGAGTGGAGCATGGCCACCTACTACCTGTTTCGAC GAGAGCGGCGAGTTCAAGCTGGCCAGCCACATGTACTGCAGCTTCTAC CCCCCGACGAGGACGAGGAGGAGGGCGACTGCGAGGAGGAGGAGT TCGAGCCCAAGCACCCAGTACGAGTACGGCACCGAGGACGACTACCAG GGCAAGCCCCCTGGAGTTTCGGCGCCACCAGCGCCGCGCTGCAGCCGA GGAGGAGCAGGAGGAGGAGTGGCTGGACGACGACAGCCAGCAGACG GTGGGCCAGCAGGACGGCAGCGAGGACAACCAGACCACCACCATCCA GACCATCGTGGAGGTGCAGCCCCAGCTGGAGATGGAGCTGACCCCGG TGGTGCAGACCATCGAGGTGAACAGCTTCAGCGGCTACCTGAAGCTGA CCGACAACGTGTACATCAAGAACGCCGACATCGTGGAGGAGGCCAAGA AGGTGAAGCCCCACCGTGGTGGTGAACGCCGCCAACGTGTACCTGAAG CACGGCGCGGGCTGGCCGGCGCCCTGAACAAGGCCACCAACAACG

	CCATGCAGGTGGAGAGCGACGACTACATCGCCACCAACGGCCCCCTGA AGGTGGGGCGGCAGCTGCGTGCTGAGCGGCCACAACCTGGCCAAGCAC TGCCTGCACGTGGTGGGCCCCAACGTGAACAAGGGCGAGGACATCCA GCTGCTGAAGAGCGCCTACGAGAACCTCAACCAGCACGAGGTGCTGCT GGCCCCCTGCTGAGCGCCGGCATCTTCGGCGCCGACCCCATCCACA GCCTGAGGGTGTGCGTGGACACCGTGAGGACCAACGTGTACCTGGCC GTGTTCCGACAAGAACCCTGTACGACAAGCTGGTGAGCAGCTTCCTGGAG ATGAAGAGCGAGAAGCAGGTGGAGCAGAAGATCGCCGAGATCCCCAA GGAGGAGGTGAAGCCCTTCATCACCGAGAGCAAGCCCAAGCGTGGAGC AGAGGAAGCAGGACGACAAGAAGATCAAGGCCCTGCGTGGAGGAGGTG ACCACCACCTTGGAGGAGACCAAGTTCCTGACCGAGAACCTGCTGCTG TACATCGACATCAACGGCAACCTGCACCCCGACAGCGCCACCCCTGGTG AGCGACATCGACATCACCTTCCTGAAGAAGGACGCCCCCTACATCGTG GGCGACGTGGTGCAGGAGGGCGCTGCTGACCGCCGTGGTGATCCCCAC CAAGAAGGCCCGCGCGCACCAACCGAGATGCTGGCCAAGGCCCTGAGGA AGGTGCCACCGACAACCTACATCACCACTACCCCGGCCAGGGCCCTGA ACGGCTACACCGTGGAGGAGGCCAAGACCGTGTGCTGAAGAAGTGCAAG AGCGCCTTCATATCCTGCCCAGCATCATCAGCAACGAGAAGCAGGAG ATCCTGGGCAACCGTGAGCTGGAACCTGAGGGAGATGCTGGCCCCACGC CGAGGAGACCAGGAAGCTGATGCCCGTGTGCGTGGAGACCAAGGCCA TTCGTGAGCACCATCCAGAGGAAGTACAAGGGCATCAAGATCCAGGAGG CGGTGGTGGACTACGGCGCCAGGTTCCTACTTCTACACCAGCAAGACCA CCGTGGCCAGCCTGATCAACACCCCTGAACGACCTGAACGAGACCCCTGG TGACCATGCCCTGGGCTACGTGACCCACGGCCCTGAACCTGGAGGAG GCCGCCAGGTACATGAGGAGCCTGAAGGTGCCCGCCACCGTGAGCGT GAGCAGCCCCGACCGCGTGACCGCCTACAACGGCTACCTGACCAAGCA GCAGCAAGACCCCCGAGGAGCACTTCATCGAGACCATCAGCCTGGCC GGCAGCTACAAGGACTGGAGCTACAGCGGCCAGAGCACCCAGCTGGG CATCGAGTTCCTGAAGAGGGGGCGACAAGAGCGTGTACTACACCAGCAA CCCCACCACCTTCCACCTGGACGGCGAGGTGATCACCTTCGACAACCT GAAGACCCCTGCTGAGCCTGAGGGAGGTGAGGACCATCAAGGTGTTCAC CACCGTGGACAACATCAACCTGCACGGAAGCGGAGCCACGAACCTCTC TCTGTTAAAGCAAGCAGGAGATGTTGAAGAAAACCCCGGGCCTAGCGG CTTCAGGAAGATGGCTTCCCCAGCGGCAAGGTGGAGGGCTGCATGG TGCAGGTGACCTGCGGCACCAACACCCCTGAACGGCCTGTGGCTGGA CGACGTGGTGTACTGCCCCAGGCACCTGATCTGCACCAAGCGAGGACA TGCTGAACCCCCAATACGAGGACCTGCTGATCAGGAAGAGCAACCCAC AATTCCTGGTGCAGGCCGGCAACGTGCAGCTGAGGGTGAAGGCCCA CAGCATGCAGAACTGCGTGCTGAAGCTGAAGGTGGACACCGCCAAACC CCAAGACCCCCAAGTACAAGTTCGTGAGGATCCAGCCCGGCCAGACC TTCAGCGTGCTGGCCTGCTACAACGGCAGCCCCAGCGGCGTGTACCA GTGCGCCATGAGGCCCAACTTCACCATCAAGGGCAGCTTCCTGAACG GCAGCTGCGGCAGCGTGGGCTTCAACATCGACTACGACTGCGTGAGC TTCGTGCTACATGCAGGAGATGGAGCTGCCGACCGGCGTGCAGGCCGG CACCGACCTGGAGGGCAACTTCTACGGCCCCCTTCGTGGACAGGCAGA CCGCCCCAGGCCGCGCGCACCGACACCACCATCACCGTGAACGTGCT GGCCTGGCTGTACGCCCGCGTGATCAACGGCGACAGGTGGTTCCTGA ACAGCTTCACCACCACCTGAACGACTTCAACCTGGTGGCCATGAAG TACAATACGAGCCCCGACCCAGGACCACGTGGACATCTTGGGCCC CCTGAGCGCCCCAGACCGGCATCGCCGTGCTGGACATGTGCGCCAGCC TGAAGGAGCTGCTGCAGAACGGCATGAACGGCAGGACCATCTGGG CAGCGCCCTGCTGGAGGACGAGTTCACCCCTTCGACGTGGTGAGGC AGTGCAGCGGCGTGACCTTCAGGGAAGCGGAGCCACGAACCTCTCT CTGTTAAAGCAAGCAGGAGATGTTGAAGAAAACCCCGGGCCTAGCGG GAGCGCACGCTTCCTGAACCCCTTCGCGGCTGAGCGGCCGAGC GCTGACCCCTGCGGCAACCGGACGAGGACCGAGCTGCTGTACAGCG CCTTCGACATCTACAACGACAAGGTGGCGGGCTTCGCAAGTTCTCTCA AGAGGAAGTGTGCGCTTCAGGACGAGGACGAGGACGAGGACGAGGACG	
--	---	--

	ATGSAAGGCTACTTCCTGGTGAAGAGGGGACACCTTCAGGAACTACGAG CACGAGGAGACCATCTACAACCTGCTGAAGGACTGCCCGCGCGGTGGGG AAGGACGACTTCTTCAAGTTCAAGGATCGACGGCGGACATGGTGGCCGAC ATCAGCAGGCAGAGGCTGACCAAGTACACCATGGCGGAGCTGGTGTAC GCCCTGAGGCACTTCGACGAGGGGCAACTGCGACACCCCTGAAGGAGAT CCTGGTGACCTACAAC TGCTGCCGAGGACCACTACTTCAACAAGAAGGA CTGGTACGACTTCGTGGAGAACCCCGACATCCTGAGGGGTGTAGGCCAA CCTGGGGCAGAGGGCTGAGGCAAGGCCCTGCTGAAGACCGTGCAGTTCT GCGACGCCATGAGGAACGCCGGCATCGTGGGGGTGCTGACCCCTGGAC AAGCAGGACCTGAACGGCAACTGGTACGACTTCGGCGACTTCATCCAG ACCAGCCCGCGGCGAGGGGCTGCCCGGTGCTGGACAGCTACTACAGCC GCTGATGCCCATCCTGACCCCTGACCAAGGGCCCTGACCGCGGAGAGCC AGGTGGAGACCGAGCTGACCAAGCCCTACATCAAGTGGGACCTGCTGA AGTACGACTTCACCGACGAGAGGGCTGAAGCTGTTCCGACAGGTACTTCA AGTATGGGACGAGACCTACCAACCCCAACTGGCTGAAGTGGCTGGAGC ACAGGTGCATCCTCCACTCCGCCAACTTCAACGTGCTGTTACGACACCG TGTTCGCCCCCACCAGCTTCGGCCCCCTGGTGAGGAAATCTTTCGTGG ACGGGGTGCCCTTCGTGGTGAGGACCGGCTACCCTTCAGGCGAGCTG GCGGTGCTGCACAACCAGGACGTGAACCTGCACAGCAGCAGGCTGAG CTTCAAGGAGCTGCTGGTGTACGCCGCCGACCCCGCCATGCACGCCG CCAGCGGCAACCTGCTGCTGGACAAGAGGACCACCTGCTTCAGCGTG GCCCGCCTGACCAAGAACGTGGCCCTTCAGACCGTGAAGCCCGGCA CTTCAACAAGGACTTGTACGACTTCGCCGTGAGCAAGGGCTTCTTCAAG GAGGGCAGCAGCGTGGAGCTGAAGCACTTCTTCTTCGCCCAGGACCG CAAGGGCGGCCATCAGCGACTACGACTACTACAGGTACAACCTGCCGAG CATGTGCGACATCAGGCGAGCTGCTGTTCCGTGGTGGAGGTGGTGGACAA CTACTTCGACTGCTACGACGGCGGCTGCATCAACGGCAACAGGTTGAT CGTGAACAACCTGGACAAGAGCGCGGGCTTCCCCCTTCAACAAGTGGG GCAAGGGCAGGCTGTACTACGACAGCATGAGCTAGGAGGACCGAGGAG CCCTGTTCCCTACACCAAGAGGAAGCTGATGCCGACCATCACCCAGAT GAACCTGAAGTACGGCATCAGCGCCAAAGAACAGGGCCAGGACCGTGC CGCGCTGAGCATCTGCAGCACCATGACCAACAGGCGAGTTCCAGCAGA AGCTGCTGAAGAGCATCGCCGCCACCAAGGGGGCGCCACCGTGGTGATC GGCAGCAGCAAGTTCTACGGCGGCTGGGACAACATGCTGAAGAGCGGTG TACAGCGACGTGGAGAACCCCACTGATGGGCTGGGACTACCCCAAG TGGGACAGGGCCATGCCCAACATGCTGAGGATCATGGCCAGCCCTGGTG CTGGCCAGGAAGTACACCCCTGCTGCGAGCCTGAGCCACAGGTTCTAG AGGCTGGCCCAACGAGTGGCGCCAGGTGCTGAGCGAGATGGTGATGTG CGCGCGCAGCCTGTAGGTGAAGCCCGGCGGCGCAGCAGCAGCGCGGAG GCCACACCGCTACGCCAACAGCGTGTTCAGCATCTGCCAGGCGCGTG AGCGGCAACGTGAAGCGGCTGCTGAGCAGCGAGGGCAACAAGATCGG CGACAAGTACCTGAGGAACCTGCAGCACAGGCTGTACGAGTGGCTGTG CAGGAACAGGGACGTGCACACCGACTTCGTGAACGAGTTCTACCCCTA CCTGAGGAAGCAGTTACCATGATGATCCTGAGCGACGACGCGGTGGT CTGCTTCAACAGGACCTACGCCAGCGAGGGCCCTGGTGGCCAGCATCAA GAACCTCAAGAGCGTGGTGTACTACCAAGAACAGGTGTTGATGAGCGAG GCCAAGTGCTGGACCGAGACCGACCTGACCAAGGGCGCCCAAGAGTT CTGAGCGCAGCAGCATGCTGGTGAAGCAGGGCGACGACTACCTGTG CCTGGCCTACCCGACCCAGCAGCATCCTGGGCGCCGGCTGCTTCC TGGACGACATCGTGAAGACCGACGGCACCCCTGATGATCGAGAGGTTCC TGAGCCTGGCCATCGACGGCTACCCCTGACCAAGCAGCGCAACGAGG AGTACGCCGAGGTGTTCCACCTGTACCTGCAGTACATCAGGAAGCTGCA CGAGGAGCTGACCGGCGACATGCTGGACATGTACAGCGTGTGCTGAG CAACGACAACACAGCAGGTACTGGGAGCCCGAGTTCTACGAGGCCAT CTACAGCCCGCAGACCGTGGTGGCACTGACTCGAGCTGGTACTGCATGC <u>ACGCAATGCTAGCTGCCCTTTCCCGTCCCTGGGTACCCCGAGTCTCCC</u> <u>CCGACCTCGGGTCCCAGGTATGCTCCACCTCCACCTGCCCGACTCA</u> <u>CCACCTCTGCTAGTTCCAGACACCTCCCAAGCACGCAGCAATGCAGC</u>	
--	---	--

	<u>TCAAAACGCTTAGCCTAGCCACACCCCCACGGGAAACAGCAGTGATT</u> <u>AACCTTTAGCAATAAACGAAAGTTTAACTAAGCTATACTAACCCAGGG</u> <u>TTGGTCAATTCGTGCCAGCCACACCTGGAGCTAGCAAAAAAAAA</u>	
6 <i>promoter,</i> <i>5'UTR and</i> <i>leader</i> <i>sequence,</i> <i>Spike</i> <i>glycoprotein</i> <i>(HexaPro-</i> <i>mutations).</i> <i>linker NSP3,</i> <i>Stop Codon,</i> <i>3'UTR and</i> <i>PolyA tail</i>	<u>CTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT</u> <u>AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGG</u> <u>CCCCGCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAAT</u> <u>GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAAT</u> <u>GGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA</u> <u>TCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC</u> <u>GCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCTACTTGGCA</u> <u>GTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCA</u> <u>CGTTCGTCTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCAATTTT</u> <u>GTATTTATTTATTTTTTAATTATTTTGTGTCAGCGATGGGGGCGGGGGGGG</u> <u>GGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGG</u> <u>GGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGGCG</u> <u>CGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGGCGGGCGGGCGGGCCC</u> <u>TATAAAAAGCGAAGCGCGCGGGCGGGGAGTCTGCTGCGCGCTGCCT</u> <u>TGCGCCCGTGCCCCGCTCCGCGCGCGGCTCGCGCGCGCGCGCGCGCG</u> <u>CTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCC</u> <u>TTCTCCTCCGGGCTGTAAATAGCGCTTGGTTTAAATGACGGCTTGTTCCT</u> <u>TTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGGCCCTTG</u> <u>TGCGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGIGTGIGTGCGTGG</u> <u>GGAGCGCGCGTGC GGCTCCGCGCTGCCCGGCGGCTGTGAGCGCTG</u> <u>CGGGCGCGCGCGGGGCTTTGTGCGCTCCGCAGTGTGCGCGAGGGG</u> <u>AGCGCGGCGCGGGGGCGGTGCCCCGCGGTGCGGGGGGGGCTGCGAG</u> <u>GGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAG</u> <u>GGGGTGTGGGCGCGCTCGGTGCGGCTGCAACCCCCCTGCAACCCCCCT</u> <u>CCCCGAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTAC</u> <u>GGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCA</u> <u>GGTGGGGGTGCGGGGCGGGGCGGGGCGGCTCGGGCGGGGAGGG</u> <u>CTCGGGGAGGGGCGCGGCGGCCCCCGAGCGCGCGGCTGTGCG</u> <u>AGGCGCGGCGAGCGCGAGCCATTGCCCTTTTATGGTAATCGTGCAGAG</u> <u>GGCGCAGGGACTTCCTTTGTCCCAATCTGTGCGGAGCCGAAATCTGG</u> <u>GAGGCGCGCGCGCACCCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGC</u> <u>GGCGCGGCGAGGAAGGAATGGGCGGGGAGGGCCTTCGTGCGTCCG</u> <u>CGCGCGCGGCTCCCTTCTCCCTCTCCAGCCTCGGGGCTGTCCGCGG</u> <u>GGGGACGGCTGCCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCCGGT</u> <u>TCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAACCATTGTCATG</u> <u>CCTTCTTCTTTTCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCT</u> <u>GTCTCATCATTTTGGCAAAGAATTGGAGAATAAACTAGTATTCTTCTGGTC</u> <u>CCACAGACTCAGAGAGAACC CGCCACCATGTTCTGTTCCTGGTGCT</u> <u>GCTGCCCCCTGGTGAGCAGCCAGTGCGTGAACCTGACCACCAGGACCC</u> <u>AGCTGCCCCCGCCTACACCAACAGCTTCACCAGGGGCGTGTA CTACC</u> <u>CCGACAAGGTGTT CAGGAGCAGCGTGCTGCACAGCACCCAGGACCTG</u> <u>TTCCTGCCCTTCTCAGCAACGTGACCTGGTTCCACGCCATCCACGTGA</u> <u>GCGGCACCAACGGCACCAAGAGGTTGACAACCCCGTGCTGCCCTTC</u> <u>AACGACGGCGTGACTTCGCCAGCACCGAGAAGAGCAACATCATCAGG</u> <u>GGCTGGATCTTCGGCACCACTGGACAGCAAGACCCAGAGCCTGCT</u> <u>GATCGTGAACAACGCCACCAACGTGGTGATCAAGGTGTGCGAGTTCCA</u> <u>GTTCTGCAACGACCCCTTCCTGGGCGTGACTACCACAAGAACAACAA</u> <u>GAGCTGGATGGAGAGCGAGTTCAGGGGTGACAGCAGCGCCAACAAC</u> <u>GCACCTTCGAGTACGTGAGCCAGCCCTTCCTGATGGACCTGGAGGGCA</u> <u>AGCAGGGCAACTTCAAGAACCTGAGGGAGTTCGTGTTCAAGAACATCG</u> <u>ACGGCTACTTCAAGATCTACAGCAAGCACACCCCATCAACCTGGTGAG</u> <u>GGACCTGCCCCAGGGCTTCAGCGCCCTGGAGCCCTGGTGAGACTGC</u> <u>CCATCGGCATCAACATCACCAGGTTCCAGACCTGCTGGCCCTGCACA</u> <u>GGAGCTACCTGACCCCCGGCGACAGCAGCAGCGGCTGGACCGCCGG</u>	144

	CGCCGCCGCTACTACGTGGGCTACCTGCAGCCCAGGACCTTCCTGCT GAAGTACAACGAGAACGGCACCATCACCGACGCCGTGGACTGCGCCCT GGACCCCTGAGCGAGACCAAGTGCACCCTGAAGAGCTTCACCGTGG AGAAGGGCATCTACCAGACCAGCAACTTCAGGGTGCAGCCCACCGAGA GCATCGTGAGGTTCCCCAACATCACCAACCTGTGCCCTTCGGCGAGG TGTTCAACGCCACCAGGTTGCCAGCGTGTACGCCTGGAACAGGAAGA GGATCAGCAACTGCGTGGCCGACTACAGCGTGCTGTACAACAGCGCCA GCTTCAGCACCTTCAAGTGCTACGGCGTGAGCCCCACCAAGCTGAACG ACCTGTGCTTCACCAACGTGTACGCCGACAGCTTCGTGATCAGGGGCG ACGAGGTGAGGCAGATCGCCCCGGCCAGACCGGCAAGATCGCCGAC TACAACTACAAGCTGCCCCGACGACTTCACCGGCTGCGTGATCGCCTGG AACAGCAACAACCTGGACAGCAAGGTGGGCGGCAACTACAACCTACCTG TACAGGCTGTTAGGAAGAGCAACCTGAAGCCCTTCGAGAGGGACATC AGCACCGAGATCTACCAGGGCCGGCAGCACCCCTGCAACGGCGTGGA GGGCTTCAACTGCTACTTCCCCCTGCAGAGCTACGGCTTCAGCCAC CAACGGCGTGGGCTACCAGCCCTACAGGGTGGTGGTGCTGAGCTTCG AGCTGCTGCACGCCCCCGCCACCGTGTGCGGCCCAAGAAGAGCACC AACCTGGTGAAGAACAAGTGCCTGAACCTTCAACTTCAACGGCCTGACC GGCACC GGCGTGCTGACCGAGAGCAACAAGAAGTTCCTGCCCTTCCA GCAGTTCGGCAGGGACATCGCCGACACCACCGACGCCGTGAGGGACC CCCAGACCCTGGAGATCCTGGACATCACCCCTGCAGCTTCGGCGGC GTGAGCGTGATCACCCCGGCACCAACACCAGCAACCAGGTGGCCGT GCTGTACCAGGACGTGAACCTGCACCGAGGTGCCCGTGCCATCCACG CCGACCAGCTGACCCACCTGGAGGGTGTACAGCACCGGCAGCAAC GTGTTCCAGACCAGGGCCGGCTGCCTGATCGGCGCCGAGCACGTGAA CAACAGCTACGAGTGCGACATCCCCATCGGCGCCGGCATCTGCGCCAG CTACCAGACCCAGACCAACAGCCCCGGCAGCGCCAGCAGCGTGGCCA GCCAGAGCATCATCGCCTACACCATGAGCCTGGGCGCCGAGAACAGCG TGGCCTACAGCAACAACAGCATCGCCATCCCCACCAACTTCACCATCAG CGTGACCACCGAGATCCTGCCCCGTGAGCATGACCAAGACCAGCGTGGA CTGCACCATGTACATCTGCGGCGACAGCACCGAGTGACAGCAACCTGCT GCTGCAGTACGGCAGCTTCTGCACCCAGCTGAACAGGGCCCTGACCG GCATCGCCGTGGAGCAGGACAAGAACACCCAGGAGGTGTTTCGCCAG GTGAAGCAGATCTACAAGACCCCCCATCAAGGACTTCGGCGGCTTC AACTTCAGCCAGATCCTGCCCCGACCCAGCAAGCCAGCAAGAGGAG CCCCATCGAGGACCTGCTGTTCAACAAGGTGACCTGGCCGACGCCG GCTTCATCAAGCAGTACGGCGACTGCCTGGGCGACATCGCCGACGG GACCTGATCTGCGCCAGAAAGTTCAACGGCCTGACCGTGCTGCCCCCC CTGCTGACCGACGAGATGATCGCCAGTACACCAGCGCCCTGCTGGCC GGCACCATCACCAGCGGCTGGACCTTCGGCGCCGGCCCGCCCTGCA GATCCCCCTTCCCATGCAGATGGCCTACAGGTTCAACGGCATCGGCGT GACCCAGAACGTGCTGTACGAGAACCAGAAGCTGATCGCCAACCAGTT CAACAGCGCCATCGGCAAGATCCAGGACAGCCTGAGCAGCACCCCA GCGCCCTGGGCAAGCTGCAGGACGTGGTGAACCAGAACGCCAGGCC CTGAACACCCTGGTGAAGCAGCTGAGCAGCAACTTCGGCGCCATCAGC AGCGTGCTGAACGACATCCTGAGCAGGCTGGACCCCCGAGGCCGA GGTGACATCGACAGGCTGATCACCGGACAGGCTGCAGAGCCTGCAGA CCTACGTGACCCAGCAGCTGATCAGGGCCGCGGAGATCAGGGCCAGC GCCAACCTGGCCGCCACCAAGATGAGCGAGTGCGTGCTGGGCCAGAG CAAGAGGGTGGACTTCTGCGGCAAGGGCTACCACCTGATGAGCTTCCC CCAGAGCGCCCCCACGGCGTGGTGTTCCTGCACGTGACCTACGTGC CCGCCCAGGAGAAGAACTTCACCACCGCCCCCGCCATCTGCCACGAC GGCAAGGCCCACTTCCCCAGGGAGGGCGTGTTCTGTGAGCAACGGCAC CCTGTTCTGTGACCCAGAGGAATTTCTACGAGCCCCAGATCATCAC CACCGACAACACCTTCGTGAGCGGCAACTGCGACGTGGTGATCGGCAT CGTGAACAACACCGTTACGACCCCTGCAGCCCGAGCTGGACGCTT CAAGGAGGAGCTGGACAAGTACTTCAAGAACCACACAGCCCGACGT GGACCTGGGCGACATCAGCGGCATCAACGCCAGCGTGGTGAACATCCA	
--	---	--

	GAAGGAGATCGACAGGCTGAACGAGGTGGCCAAGAACCTGAACGAGA GCCTGATCGACCTGCAGGAGCTGGGCAAGTACGAGCAGTACATCAAGT GGCCCTGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCATCG TGATGGTGACCATCATGCTGTGCTGCATGACCAGCTGCTGCAGCTGCC GAAGGGCTGCTGCAGCTGCGGCAGCTGCTGCAAGTTCGACGAGGACG ACAGCGAGCCCGTGCTGAAGGGCGTGAAGCTGCACTACACCGGAAGC GGAGCCACGAACCTCTCTCTGTTAAAGCAAGCAGGAGATGTTGAAGAAA ACCCCGGGCCTGCCCCACCAAGGTGACCTTCGGCGACGACACCGTG ATCGAGGTGCAGGGCTACAAGAGCGTGAACATCACCTTCGAGCTGGAC GAGAGGATCGACAAGGTGCTGAACGAGAAGTGCAGCGCTACACCGT GGAGCTGGGCACCGAGGTGAACGAGTTCGGCTGCGTGGTGGCCGACG CCGTGATCAAGACCTTGCAGCCCGTGAGCGAGCTGCTGACCCCGCTG GGCATCGACCTGGACGAGTGGAGCATGGCCACCTACTACCTGTTCCGAC GAGAGCGGGGAGTTCAAGCTGGCCAGCCACATGTACTGACGCTTCTAC CCCCCGACGAGGACGAGGAGGAGGGCGACTGCCAGGAGGAGGAGT TCGAGCCGACGACCCAGTACGAGTACGGCACCGAGGACGACTACCAG GGCAAGCCCTGCGAGTTCCGCCCGCACCGCGCGCCCTGCGAGCCGA GGAGGAGCAGGAGGAGGACTGGCTGGACGACGACAGCCAGCAGACC GTGGGCCAGCAGGACGGCAGCGAGGACAACCAGACCACCACTCCA GACCATCGTGGAGGTGCAGCCCCAGCTGGAGATGGAGCTGACCCCG TGGTGCAGACCATCGAGGTGAACAGCTTCAGCGGCTACCTGAAGCTGA CCGACAACGTGTACATCAAGAACCGGACATCGTGGAGGAGGCCAAGA AGGTGAAGCCACCGTGGTGGTGAACCGCCGCAACGTGTACCTGAAG CACCGCGCGCGCGTGGCCGCGCGCTGAACAAGGCCACCAACAACG CCATGCAGGTGGAGAGCGACGACTACATCGCCACCAACGGCGCGCTGA AGGTGGGCGCGCAGCTGCGTGTGAGCGCGCCACAACCTGGCCAAAGCAC TGCTGCACGTGGTGGGCCCCAACGTGAACAAGGGCGAGGACATCCA GCTGCTGAAGAGCGCCTACGAGAACCTCAACCAGCAGGAGGTGCTGCT GGCCCCCTGCTGAGCGCGCGCATCTTCGGCGCGGACCCCATCCACA GCCTGAGGGTGTGCGTGGACACCGTGAGGACCAACGTGTACCTGGCC GTGTTCCGACAAGAACCCTGTACGACAAGCTGGTGGAGCAGCTTCCTGGAG ATGAAGAGCGAGAAGCAGGTGGAGCAGAAGATCGCCGAGATCCCAA GGAGGAGGTGAAGCCCTTCATCACCGAGAGCAAGCCGAGCTGGAGC AGAGGAAGCAGGACGACAAGAAGATCAAGGCCCTGCGTGGAGGAGGTG ACCACCACCTGGAGGAGACCAAGTTCTTGACCGAGAACCTGCTGCTG TACATCGACATCAACGGCAACCTGCACCCCGACAGCGCCACCTGGTG AGCGACATCGACATCACCTTCCTGAAGAAGGACGCCCCCTACATCGTG GGCGACGTGGTGCAGGAGGGCGTGTGACCGCGCGTGGTGATCCCCAC CAAGAAGGCGCGCGCACCAACCGAGATGCTGGCCAAGGCCCTGAGGA AGGTGCCACCGACAACCTACATCACCACTACCCCGGCCAGGGCCTGA ACGGCTACACCGTGGAGGAGGCCAAGACCGTGTGAAGAAGTGCAAG AGCGCCTTCATACATCCTGCCAGCATCATCAGCAACGAGAAGCAGGAG ATCCTGGGCACCGTGAGCTGGAACCTGAGGGAGATGCTGGCCGACGC CGAGGAGACCAGGAAGCTGATGCCCGTGTGCGTGGAGACCAAGGCCA TCGTGAGCACCATCCAGAGGAAGTACAAGGGCATCAAGATCCAGGAGG CGGTGGTGGACTACGGCGCCAGGTTCATCTTACACCAAGCAAGACCA CCGTGGCCAGCCTGATCAACACCTGAACGACCTGAACGAGACCCCTGG TGACCATGCCCTGGGCTACGTGACCCACGGCCTGAACCTGGAGGAG GCCGCCAGGTACATGAGGAGCCTGAAGGTGCCCGCCACCGTGAGCGT GAGCAGCCCCGACGCGCTGACCGCCTACAACGGCTACCTGACCAAGCA GCAGCAAGACCCCCGAGGAGCACTTCATCGAGACCATCAGCCTGGCC GCCAGCTACAAGGACTGGAGCTACAGCGGCCAGAGCACCCAGCTGGG CATCGAGTTCTGAAGAGGGGGGACAAGAGCGTGTACTACACCAGCAA CCCCACCACCTTCACCTGGACGGCGAGGTGATCACCTTCGACAACCT GAAGACCTGCTGAGCCTGAGGGAGGTGAGGACCATCAAGGTGTTCAC CACCGTGGACAACATCAACCTGCACCTGACTCGAGCTGGTACTGCATGC ACGCAATGCTAGCTGCCCTTTCCCGTCCCTGGGTACCCCGAGTCTCCC CCGACCTCGGGTCCCAGGTATGCTCCACCTCCACCTGCCCCACTCA	
--	--	--

	<u>CCACCTCTGCTAGTTCAGACACCTCCCAAGCACGCAGCAATGCAGC</u> <u>TCAAAACGCTTAGCCTAGCCACACCCCCACGGGAAACAGCAGTGATT</u> <u>AACCTTTAGCAATAAACGAAAGTTTAACTAAGCTATACTAACCCAGGG</u> <u>TTGGTCAATTCGTGCCAGCCACACCTGGAGCTAGCAAAAAAAA</u>	
7 <i>promoter,</i> 5'UTR and leader sequence, Spike glycoprotein (HexaPro- <i>mutations</i>), linker <u>NSP5</u>, <u>Stop Codon,</u> 3'UTR and <u>PolyA tail</u>	<u>CTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT</u> <u>AGTTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGG</u> <u>CCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAAT</u> <u>GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAAT</u> <u>GGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA</u> <u>TCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC</u> <u>GCCTGGCATTTATGCCAGTACATGACCTTATGGGACTTTCCTACTTTGGCA</u> <u>GTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCA</u> <u>CGTTCTGCTTCACCTCTCCCCATCTCCCCCCCCCTCCCCACCCCCAATTTT</u> <u>GTATTTATTTATTTTTAATTATTTTGTGCAGCGATGGGGCGGGGGGGG</u> <u>GGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGG</u> <u>GGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCG</u> <u>CGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGGCGGGCGGCGG</u> <u>TATAAAAGCGAAGCGCGCGGGCGGGCGGGAGTCGCTGCGCGCTGCCCT</u> <u>TCGCCCCGTGCCCGCTCCGCGCGCGCTCGCGCGCGCGCGCGCGG</u> <u>CTCTGACTGACCGCGTTACTCCACAGGTGAGCGGGCGGGACGGGCC</u> <u>TTCTCCTCCGGGCTGTAATTAGCGCTTGGTTAATGACGGCTTGTTCTT</u> <u>TTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGGCCCTTG</u> <u>TGCGGGGGGAGCGGGCTCGGGGGGTGCGTGCGTGTTGTGTGCGTGG</u> <u>GGAGCGCGCGTGCGGGCTCCGCGCTGCCCGGGCGGCTGTGAGCGCTG</u> <u>CGGGCGCGGGCGCGGGGCTTTGTGCGCTCCGCAAGTGTGCGCGAGGGG</u> <u>AGCGCGGCGGGGGCGGTGCCCGCGGTGCGGGGGGGGCTGCGAG</u> <u>GGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGGTGAGCAG</u> <u>GGGGTGTGGGCGCGTGGGTGCGGGCTGCAACCCCCCTGCAACCCCCCT</u> <u>CCCCGAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTAC</u> <u>GGGGCGTGCGCGGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCA</u> <u>GGTGGGGGTGCCGGGCGGGCGGGGCGGCTCGGGCGGGGAGGG</u> <u>CTCGGGGGAGGGGGCGCGGGCGGGCCCCGGAGCGCGCGGGCTGTGG</u> <u>AGGCGCGGCGAGCCGCGAGCCATTGCCTTTTATGGTAATCGTGCGAGAG</u> <u>GGCGCAGGGACTTCCTTTGTCCCAATCTGTGCGGAGCCGAAATCTGG</u> <u>GAGGCGCGCGCGCACCCCCCTCTAGCGGGCGGGGGCGAAGCGGTGC</u> <u>GGCGCCGGCAGGAAGGAAATGGCGGGGAGGGGCTTCGTGCTGCG</u> <u>CGCGCGCGGTCCCCCTTCCTCCTCTCCAGCCTCGGGGGCTGTCCGCGG</u> <u>GGGGACGGCTGCCCTTCGGGGGGGACGGGGCAGGGCGGGGTTTCGGCT</u> <u>TCTGGCGTGTGACCGGGCGGCTCTAGAGCCTCTGCTAACCATGTTTCATG</u> <u>CCTTCTTTCTTTCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCT</u> <u>GTCTCATCATTTTGGCAAAGAATTGGAGAATAAACTAGTATTCTTCTGGTC</u> <u>CCACAGACTCAGAGAGAACCCGCCACCATGTTCTGTGTTCTGGTGCT</u> <u>GCTGCCCCCTGGTGAGCAGCCAGTGCGTGAACCTGACCACCAGGACCC</u> <u>AGCTGCCCCCGCTACACCAACAGCTTACACAGGGGCGTGTACTACC</u> <u>CCGACAAGGTGTTACAGGAGCAGCGTGTGCACAGCACCCAGGACCTG</u> <u>TTCTGCCCCCTTTCAGCAACGTGACCTGGTTCCACGCCATCCACGTGA</u> <u>GCGGCACCAACGGCACCAAGAGGTTGACAACCCCGTGTGCCCTTC</u> <u>AACGACGGCGTGACTTCGCCAGCACCGAGAAGAGCAACATCATCAGG</u> <u>GGCTGGATCTTCGGCACCAACCCTGGACAGCAAGACCCAGAGCCTGCT</u> <u>GATCGTGAACAACGCCACCAACGTGGTGATCAAGGTGTGCGAGTTCCA</u> <u>GTTCTGCAACGACCCCTTCCTGGGCGTGTAACCACAAGAACAACAA</u> <u>GAGCTGGATGGAGAGCGAGTTACAGGTGTACAGCAGCGCCAACAAC</u> <u>GCACCTTCGAGTACGTGAGCCAGCCCTTCCTGATGGACCTGGAGGGCA</u> <u>AGCAGGGCAACTTCAAGAACCTGAGGGAGTTCGTGTTCAAGAACATCG</u> <u>ACGGCTACTTCAAGATCTACAGCAAGCACACCCCATCAACCTGGTGAG</u> <u>GGACCTGCCCCAGGGCTTCAGCGCCCTGGAGCCCTGGTGACCTGC</u> <u>CCATCGGCATCAACATCACCAGGTTCCAGACCCTGCTGGCCCTGCACA</u>	145

	GGAGCTACCTGACCCCCGGCGACAGCAGCAGCGGCTGGACCGCCGG CGCCGCCGCTACTACGTGGGCTACCTGCAGCCCAGGACCTTCCTGCT GAAGTACAACGAGAACGGCACCATACCGACGCCGTGGACTGCGCCCT GGACCCCCTGAGCGAGACCAAGTGACCCCTGAAGAGCTTCACCGTGG AGAAGGGCATCTACCAGACCAGCAACTTCAGGGTGACGCCACCGAGA GCATCGTGAGGTTCCCCAACATCACCACCTGTGCCCTTCGGCGAGG TGTTCAACGCCACCAGGTTCCGCCAGCGGTACGCCTGGAACAGGAAGA GGATCAGCAACTGCGTGGCCGACTACAGCGTGCTGTACAACAGCGCCA GCTTCAGCACCTTCAAGTGCTACGGCGTGAGCCCCACCAAGCTGAACG ACCTGTGCTTCACCAACGTGTACGCCGACAGCTTCGTGATCAGGGGCG ACGAGGTGAGGCAGATCGCCCCCGGCCAGACCGGCAAGATCGCCGAC TAACTACAAGCTGCCCGACGACTTCACCGGCTGCGTGATCGCCTGG AACAGCAACAACCTGGACAGCAAGGTGGCGGGCAACTACAACCTACCTG TACAGGCTGTTCAAGGAAGAGCAACCTGAAGCCCTTCGAGAGGGACATC AGCACCGAGATCTACCAGGCCGGCAGCACCCCTGCAACGGCGTGGA GGGCTTCAACTGCTACTTCCCCCTGCAGAGCTACGGCTTCCAGCCAC CAACGGCGTGCGGCTACCAGCCCTACAGGGTGGTGGTGCTGAGCTTCG AGCTGCTGCACGCCCCCGCCACCGTGTGCGGCCCCAAGAAGAGCACC AACCTGGTGAAGAACAAGTGCGTGAACCTTCAACTTCAACGGCCTGACC GGCACCGGCGTGCTGACCGAGAGCAACAAGAAGTTCCTGCCCTTCCA GCAGTTCGGCAGGGACATCGCCGACACCACCGACGCCGTGAGGGACC CCCAGACCCTGGAGATCCTGGACATCACCCCTGCAGCTTCGGCGGC GTGAGCGTGATCACCCCGGCACCAACACCAGCAACCAAGTGGCCGT GCTGTACCAGGACGTGAACCTGCACCGAGGTGCCCGTGCCCATCCACG CCGACCAGCTGACCCCCACCTGGAGGGTGACAGCACCGGCAGCAAC GTGTTCCAGACCAGGGCCGGCTGCCTGATCGGCGCCGAGCACGTGAA CAACAGCTACGAGTGCGACATCCCCATCGGCGCCGGCATCTGCGCCAG CTACCAGACCCAGACCAACAGCCCCGGCAGCGCCAGCAGCGTGGCCA GCCAGAGCATCATCGCCTACACCATGAGCCTGGGCGCCGAGAACAGCG TGGCCTACAGCAACAACAGCATCGCCATCCCCACCAACTTCAACATCAG CGTGACCACCGAGATCCTGCCCGTGAGCATGACCAAGACCAGCGTGGA CTGCACCATGTACATCTGCGGCGACAGCACCGAGTGACAGCAACCTGCT GCTGCAGTACGGCAGCTTCTGCACCCAGCTGAACAGGGCCCTGACCG GCATCGCCGTGGAGCAGGACAAGAACACCCAGGAGGTGTTCCGCCAG GTGAAGCAGATCTACAAGACCCCCCCCATCAAGGACTTCGGCGGCTTC AACTTCAGCCAGATCCTGCCCGACCCCAGCAAGCCCAGCAAGAGGAG CCCCATCGAGGACCTGCTGTTCAACAAGGTGACCTGGCCGACGCCG GCTTCATCAAGCAGTACGGCGACTGCCTGGGCGACATCGCCGCCAGG GACCTGATCTGCGCCCAAGATTCAACGGCCTGACCGTGCTGCCCCC CTGCTGACCGACGAGATGATCGCCAGTACACAGCGCCCTGCTGGCC GGCACCATCACAGCGGCTGGACCTTCGGCGCCGGCCCGGCCCTGCA GATCCCCTTCCCATGACAGATGGCCTACAGGTTCAACGGCATCGGCGT GACCCAGAACGTGCTGTACGAGAACCAGAAGCTGATCGCCAACCAAGT CAACAGCGCCATCGGCAAGATCCAGGACAGCCTGAGCAGCACCCCA GCGCCCTGGGCAAGCTGCAGGACGTGGTGAACAGAACGCCAGGCC CTGAACACCCTGGTGAAGCAGCTGAGCAGCAACTTCGGCGCCATCAGC AGCGTGCTGAACGACATCCTGAGCAGGCTGGACCCCCCGAGGCCGA GGTGACATCGACAGGCTGATCACCGGCAGGCTGCAGAGCCTGCAGA CCTACGTGACCCAGCAGCTGATCAGGGCCGCCGAGATCAGGGCCAGC GCCAACCTGGCCGCCACCAAGATGAGCGAGTGCGTGCTGGGCCAGAG CAAGAGGGTGGACTTCTGCGGCAAGGGCTACCACCTGATGAGCTTCCC CCAGAGCGCCCCCACGGCGTGGTGTTCCTGCACGTGACCTACGTGC CCGCCCAGGAGAAGAACTTACCACCGCCCCCGCCATCTGCCACGAC GGCAAGGCCCCACTTCCCCAGGGAGGGCGTGTTCTGTGAGCAACGGCAC CCTGTTCTGTGACCCAGAGGAATCTACGAGCCCCAGATCATCAC CAGGACAACACCTTCGTGAGCGGCAACTGCGACGTGGTGATCGGCAT CGTGAACAACACCGTGTACGACCCCTGCAGCCCCGAGCTGGACAGCTT CAAGGAGGAGCTGGACAAGTACTTCAAGAACCACACCAGCCCCGACGT	
--	--	--

	GGACCTGGGCGACATCAGCGGCATCAACGCCAGCGTGGTGAACATCCA GAAGGAGATCGACAGGCTGAACGAGGTGGCCAAGAACCTGAACGAGA GCCTGATCGACCTGCAGGAGCTGGGCAAGTACGAGCAGTACATCAAGT GGCCCTGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCATCG TGATGGTGACCATCATGCTGTGCTGCATGACCAGCTGCTGCAGCTGCCT GAAGGGCTGCTGCAGCTGCGGCAGCTGCTGCAAGTTCGACGAGGACG ACAGCGAGCCCGTGTGAAGGGCGTGAAGCTGCACTACACCGGAAGC GGAGCCACGAACCTTCTCTGTAAAGCAAGCAGGAGATGTTGAAGAAA ACCCCGGGCCTAGCGGCTTCAGGAAGATGGCCTTCCCCAGCGGCAAG CTGGAGGGCTGCATGGTGCAGCTGACCTGCGGCACCAACACCTGA ACGGCCTGTGGCTGGAGGACGCTGGTGTACTGCCCCAGGCACGTGATC TGCACCAAGCGAGGACATGCTGAACCCCAACTACGAGGACCTGCTGAT CAGGAAGAGCAACCACTTCTGGTGCAGGCGGGCAACGTGCAG CTGAGGGTGTATCGGCCACAGCATGCAGAAGTGGCTGCTGAAGCTGAA GGTGGACACCGCCCAACCCCAAGACCCCAAGTACAAGTTCGTGAGGA TCCAGCCCGGCCAGACCTTCAGCGTGTGGCCTGCTACAACGGCAGC CCCAGCGGGCTGTACCACTGCGCCATGAGGGCCCAACTTCACCATCAA GGGCAGCTTCTGAACGGCAGCTGCGGCAGCGTGGGCTTCAACATCG ACTACGACTGCGTGAGCTTCTGCTACATGCACCACATGGAGCTGCCCA CCGGCGTGCACGCGCGCACCGACCTGGAGGGCAACTTCTACGGGCC CTTCGTGGACAGGCAGACCGCCAGGCGCGCGGCACCGACACCAAC ATCACCGTGAACGTGCTGGCCTGGCTGTACGCGCGCGTGTCAACGG CGACAGGTGGTTCCTGAACAGGTTCACCAACACCTTGAACGACTTCA ACCTGGTGGCCATGAAGTACAACACGAGGCCCCTGACCCAGGACCA GTGGACATCTTGGCCCCCTGAGCGCCAGACCGGCATCGCCGTGCT GGACATGTGCGCCAGCCTGAAGGAGCTGCTGCAGAACCGCATGAAC GGCAGGACCATCCTGGGCAGCGCCCTGCTGGAGGACGAGTTCACCC CCTTCGACGTGGTGAGGCAGTGCAGCGCGGTGACCTTCCAGTGACTC GAGCTGGTACTGCATGCACGCAATGCTAGCTGCCCCCTTCCCGTCTG GGTACCCCGAGTCTCCCCGACCTCGGGTCCAGGTATGCTCCCACC TCCACCTGCCCCACTCACCACCTCTGCTAGTTCAGACACCTCCCAAG CAGCGAGCAATGCAGCTCAAAACGCTTAGCCTAGCCACACCCCCAGC GGAAACAGCAGTGATTAACCTTTAGCAATAAACGAAAGTTTAAGTAA CTATACTAACCCAGGGTTGGTCAATTTCTGTGCCAGCCACACCCCTGGA GCTAGCAAAAAAAAAA	
8 <i>promoter, 5'UTR and leader sequence, Spike glycoprotein (HexaPro- mutations), linker.  Stop Codon, 3'UTR and PolyA tail</i>	CTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGG CCCGCCTGGCTGACCGCCCAACGACCCCGCCCATTTGACGTCAATAAT GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAAT GGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA TCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC GCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCTACTTGGCA GTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCA CGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCAATTTT GTATTTATTTATTTTTAATTATTTGTGCAGCGATGGGGGCGGGGGGGG GGGGGGGGCGCGCGCCAGGCGGGGGCGGGGGCGGGGCGAGGGGGCGG GGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGGCG CGCTCCGAAAGTTTCTTTTATGGCGAGGCGGGCGGGCGGGCGGGCCCC TATAAAAAGCGAAGCGCGCGGGCGGGGCGGGAGTCGCTGCGCGCTGCCCT TCGCCCCGTGCCCCGCTCCGCGCGCGCTCGGGCGGGCGGGCGGGCGG CTCTGACTGACCGCGTTACTCCCAAGGTGAGCGGGCGGGGACGGGCC TTCTCTCCGGGCTGTAAATAGCGCTTGGTTTAAATGACGGCTTGTTCTT TTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGGCCCTTTG TGCGGGGGGAGCGGCTCGGGGGGTGCGTGCCTGTGTGTGTGCGTGG GGAGCGCGCGGTGCGGCTCCGCGCTGCCCGGGCGGCTGTGAGCGCTG CGGGCGCGGGCGGGGGCTTTGTGCGCTCCGCAAGTGTGCGCGAGGGG AGCGCGCGGGGGGGGGTGGCCCCGCGGTGCGGGGGGGGCTGCGAG	146

	<u>GGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAG</u> <u>GGGGTGTGGGCGCGTCTGGTCTGGGCTGCAACCCCCCTGCACCCCCCT</u> <u>CCCCGAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTAC</u> <u>GGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCA</u> <u>GGTGGGGGTGCGGGGCGGGGCGGGGCGGCTCGGGGCGGGGAGGG</u> <u>CTCGGGGAGGGGCGGGGCGGGGCGGGGCGGAGCGCGGGCGGCTGTG</u> <u>AGGCGCGGCGAGCCGCGAGCCATTGCTTTTATGGTAATCGTGCGAGAG</u> <u>GGCGCAGGGACTTCCTTTGTCCCAATCTGTGCGGAGCGCAATCTGG</u> <u>GAGGCGCGCGCGCAACCCCCCTCTAGCGGGGCGCGGGGCGAAGCGGTGC</u> <u>GGCGCGGCGAGGAAGGAAATGGGCGGGGAGGGGCTTCGTGCGTCCG</u> <u>CGGCGCGCGCTCCCTTCTCCCTCTCCAGCCTCGGGGCTGTCCGCGG</u> <u>GGGGACGGCTGCGCTTCGGGGGGGACGGGGCAGGGCGGGGTTCCGGCT</u> <u>TCTGGCGTGTGACCGCGGCTCTAGAGCCTCTGCTAACCATGTTTATG</u> <u>CCTTCTTCTTTTCTTCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCT</u> <u>GTCTCATCATTTTGGCAAAGAATTGGAGAATAAACTAGTATTCTTCTGGT</u> <u>CCCACAGACTCAGAGAGAACCCGCCACCATGTTCTGTGTTCTGGTGCT</u> <u>GCTGCCCCCTGGTGAGCAGCCAGTGCGTGAACCTGACCACCAGGACCC</u> <u>AGTGCCCCCGCCTACACCAACAGCTTACCAGGGGCGTGACTACC</u> <u>CCGACAAGGTGTTACGAGCAGCGTGCTGCACAGCACCCAGGACCTG</u> <u>TTCTGCCCTTCTTACGAACGTGACCTGTTCCACGCCATCCACGTGA</u> <u>GCGGCACCAACGGCACCAAGAGTTTCGACAACCCCGTGCTGCCCTTC</u> <u>AACGACGGCGTGACTTCGCCAGCACCGAGAAGAGCAACATCATCAGG</u> <u>GGCTGGATCTTCGGCACCAACCCTGGACAGCAAGACCCAGAGCCTGCT</u> <u>GATCGTGAACAACGCCACCAACGTGGTGATCAAGGTGTGCGAGTTCCA</u> <u>GTTCTGCAACGACCCCTTCTGGGCGTGACTACCACAAGAACAACAA</u> <u>GAGCTGGATGGAGAGCGAGTTACAGGGTGACAGCAGCGCCAACAACT</u> <u>GCACCTTCGAGTACGTGAGCCAGCCCTTCTGATGGACCTGGAGGGCA</u> <u>AGCAGGGCAACTTCAAGAACCTGAGGGAGTTCTGTTTCAAGAACATCG</u> <u>ACGGCTACTTCAAGATCTACAGCAAGCACACCCCCATCAACCTGGTGAG</u> <u>GGACCTGCCCCAGGGCTTCAGCGCCCTGGAGCCCCTGGTGACCTGC</u> <u>CCATCGGCATCAACATCACCAGGTTCCAGACCCTGCTGGCCCTGCACA</u> <u>GGAGCTACCTGACCCCCGGCGACAGCAGCAGCGGCTGGACCGCCGG</u> <u>CGCCGCGCCTACTACGTGGGCTACCTGCAGCCCAGGACCTTCTGTCT</u> <u>GAAGTACAACGAGAACGGCACCATCACCGACGCCGTGGACTGCGCCCT</u> <u>GGACCCCCTGAGCGAGACCAAGTGACCCCTGAAGAGCTTACCGTGG</u> <u>AGAAGGGCATCTACCAGACCAGCAACTTCAGGGTGACGCCACCGAGA</u> <u>GCATCGTGAGGTTCCCCAACATCACCAACCTGTGCCCTTCGGCGAGG</u> <u>TGTTCAACGCCACCAGGTTCCGCCAGCGGTACGCCTGGAACAGGAAGA</u> <u>GGATCAGCAACTGCGTGGCCGACTACAGCGTGCTGTACAACAGCGCCA</u> <u>GCTTCAGCACCTTCAAGTGCTACGGCGTGAGCCCCACCAAGCTGAACG</u> <u>ACCTGTGCTTCACCAACGTGTACGCCGACAGCTTCTGTATCAGGGGCG</u> <u>ACGAGGTGAGGCAGATCGCCCCCGGCCAGACCGGCAAGATCGCCGAC</u> <u>TACAACTACAAGCTGCCCCGACGACTTACCAGGCTGCGTGATCGCCTGG</u> <u>AACAGCAACAACCTGGACAGCAAGGTGGGCGGCAACTACAACCTACCTG</u> <u>TACAGGCTGTTACAGGAAGAGCAACCTGAAGCCCTTCGAGAGGGACATC</u> <u>AGCACCGAGATCTACCAGGCCGGCAGCACCCCTGCAACGGCGTGGA</u> <u>GGGCTTCAACTGCTACTTCCCCCTGCAGAGCTACGGCTTCCAGCCAC</u> <u>CAACGGCGTGGGCTACCAGCCCTACAGGGTGGTGGTGCTGAGCTTCG</u> <u>AGCTGCTGCACGCCCCCGCCACCGTGTCGGCCCCAAGAAGAGCACC</u> <u>AACCTGGTGAAGAACAAGTGCGTGAACCTTCAACTTCAACGGCCCTGACC</u> <u>GGCACCGGCGTGCTGACCGAGAGCAACAAGAAGTTCTGCCCTTCCA</u> <u>GCAGTTCCGGCAGGGACATCGCCGACACCACCGACGCCGTGAGGGACC</u> <u>CCCAGACCCTGGAGATCTGGACATCACCCCTGCAGCTTCGGCGGC</u> <u>GTGAGCGTGATCACCCCGGCACCAACACCAGCAACAGGTGGCCGT</u> <u>GCTGTACCAGGACGTGAACCTGCACCGAGGTGCCCGTGGCCATCCACG</u> <u>CCGACCAGCTGACCCCACTGGAGGGTGACAGCACCAGCGCAGCAAC</u> <u>GTGTTCCAGACCAGGGCCGGCTGCCTGATCGGCGCCGAGCAGCTGAA</u> <u>CAACAGCTACGAGTGCGACATCCCCATCGGCGCCGGCATCTGCGCCAG</u>	
--	--	--

	CTACCAGACCCAGACCAACAGCCCCGGCAGCGCCAGCAGCGTGGCCA GCCAGAGCATCATCGCCTACACCATGAGCCTGGGCGCCGAGAACAGCG TGGCCTACAGCAACAACAGCATCGCCATCCCCACCAACTTCAACATCAG CGTGACCACCGAGATCCTGCCCGTGGCAGTACCAAGACCAGCGTGA CTGCACCATGTACATCTGCGGCGACAGCACCGAGTGCAGCAACCTGCT GCTGCAGTACGGCAGCTTCTGCACCCAGCTGAACAGGGCCCTGACCG GCATCGCCGTGGAGCAGGACAAGAACACCCAGGAGGTGTTCCGCCAG GTGAAGCAGATCTACAAGACCCCCCATCAAGGACTTCGGCGGCTTC AACTTCAGCCAGATCCTGCCCGACCCAGCAAGCCAGCAAGAGGAG CCCATCGAGGACCTGCTGTTCAACAAGGTGACCCTGGCCGACGCCG GCTTCATCAAGCAGTACGGCGACTGCCTGGGCGACATCGCCGCCAGG GACCTGATCTGCGCCAGAAAGTTCAACGGCCTGACCGTGCTGCCCCCC CTGCTGACCGACGAGATGATCGCCAGTACACCAGCGCCCTGCTGGCC GGCACCATACCAGCGGCTGGACCTTCGGCGCCGGCCCGCGCTGCA GATCCCCCTTCGCATGCAGATGGCCTACAGGTTCAACGGCATCGGCGT GACCCAGAACGTGCTGTACGAGAACCAGAGCTGATCGCCAACCAAGTT CAACAGCGCCATCGGCAAGATCCAGGACAGCCTGAGCAGCACCCCGCA GCGCCCTGGGCAAGCTGCAGGACGTGGTGAACCAGAACGCCAGGCC CTGAACACCCTGGTGAAGCAGCTGAGCAGCAACTTCGGCGCCATCAGC AGCGTGCTGAACGACATCCTGAGCAGGCTGGACCCCCGAGGCCGA GGTGCAGATCGACAGGCTGATCACCAGCAGGCTGCAGAGCCTGCAGA CCTACGTGACCCAGCAGCTGATCAGGGCCGCCGAGATCAGGGCCAGC GCCAACCTGGCGCCACCAAGATGAGCGAGTGCCTGCTGGGCCAGAG CAAGAGGGTGGACTTCTGCGGCAAGGGCTACCACCTGATGAGCTTCCC CCAGAGCGCCCCCACGGCGTGGTGTTCCTGCACGTGACCTACGTGC CCGCCCAGGAGAAGAACTTACCACCGCCCCCGCCATCTGCCACGAC GGCAAGGCCCCACTTCCCCAGGGAGGGCGTGTTCGTGAGCAACGGCAC CCACTGGTTTCGTGACCCAGAGGAACTTCTACGAGCCCCAGATCATCAC CACCGACAACACCTTCGTGAGCGGCAACTGCGACGTGGTGATCGGCAT CGTGAACAACACCGTGTACGACCCCTGCAGCCCGAGCTGGACAGCTT CAAGGAGGAGCTGGACAAGTACTTCAAGAACCACACCAGCCCCGACGT GGACCTGGGCGACATCAGCGGCATCAACGCCAGCGTGGTGAACATCCA GAAGGAGATCGACAGGCTGAACGAGGTGGCCAAGAACCCTGAACGAGA GCCTGATCGACCTGCAGGAGCTGGGCAAGTACGAGCAGTACATCAAGT GGCCCTGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCATCG TGATGGTGACCATCATGCTGTGCTGCATGACCAGCTGCTGCAGCTGCCT GAAGGGCTGCTGAGCTGCGGCAGCTGCTGCAAGTTTCAGCAGGACG ACAGCGAGCCCGTGTGAAGGGCGTGAAGCTGCACTACACCGGAAGC GGAGCCACGAACCTTCTCTCTGTTAAAGCAAGCAGGAGATGTTGAAGAA ACCCCGGGCCTAGGGCGGACGCCAGAGCTTCTGTAACAGGGTGTGTG GGCTGAGCGCGCGGAGGCTGAGCGCCTGCGGCAAGCGGACACGAGCA CGACGTGGTGTACAGGGCCTTCCACATCTACAACGACAAGGTGGCCG GCTTCGCCAAGTTCTTGAAGACCAACTGCTGCAGGTTCCAGGAGAAAG ACGAGGACGACAAGCTGATCGACAGCTACTTCGTGGTGAAGAGGGCAG CCTTCAGCAACTACGAGCAGGACGAGACCATCTACAAGCTGCTGAAGC ACTGGCGCGCGGTGGGCAAGGACGAACTTCTTCAAGTTGAGGATGAGG GGGACATGGTGCCCGACATCAGGACGGCAGAGGCTGACCAAGTACACCA TGGCGGACCTGGTGTACGGCCTGAGGCACTTCGACGAGGGGCAACTGG GACACGCTGAAGGAGATCCTGGTGACCTACAAGTCTGCGACGACGAG TACTTCAACAAGAAGGACTGGTACGACTTCGTGGAGAACCCCGACATCC TGAGGGTGTACGCCAAGCTGGGCGAGAGGGTGAGGCAAGGCCCTGCTG AAGACCGTGCAGTTCTGCGACGGCATGAGGAACGCCCGCATCCTGGC CGTGCTGACCGTGGACAACGAGGACGTGAACGGCAACTGGTACGACTT CGGCGACTTCATCCAGACCACCCCGCAGCGCGCTGCCCGTGGTGG ACAGCTACTACAGCCTGCTGATGCCCATCTGACCGTGAACAGGGCCG TGACCGCCGAGAGCGCAGTGGACAGCGACCTGACCAAGCCCTACATCA AGTGGGACCTGCTGAAGTACGACTTCACCGAGGAGAGGCTGAAGCTGT TGCAGAGGTACTTCAAGTGTGGGAGGAGAGGTACCAAGCAAGTGGG	
--	---	--

	GAGCTGSCCTGGAGGAGAGGTGCATCCTGCACTGCGGCAAGCTGAAG TGCTGTTTCAGCACCCGTGTTCCCGCCGACCAGCTTCGGGCCCGCTGGTGA GGAAGATCTTCGTGGACGGCGTGGCGCTTCGTGGTGAGGACCGCGCTAGC ACTTCAGGGGAGCTGGCGGTGGTGCACAACCAGGACGTGAACCTGCGAC AGCAGCAGGCTGAGCTTCAAGGAGCTGCTGGGTGTACGCGCGCCGACCC GCGGATGCAGGCGCGCCAGCGGCAAGCTGCTGCTGGACAAGAGGAGCA CCTGCTTCAGCGTGGCGCGCCCTGAGCAACACCTGGCCCTTCAGAGCC GTGAAGCGCGGCAACTTCAACAAGGACTTCTACGACTTCGCGCTGAGG AAGGGCTTCTTCAAGGAGGGGAGCAGCGGTGGAGCTGAAGCACTTCTTC TTCGCCAGGACGGCAACGCCGCGCATGAGCGACTACGACTACTACAGG TACAACCTGCCCACTATGTGCGGACATCAGGCGAGCTGCTGTTGCTGGT GAGGTGCTGGACAAGTACTTCGACTGCTACGACGCGCGCTGCTCATCAAG GCGAAGCAGGTGATGCTGAACAACCTGGACAAGAGCGCGCGCTTCGCG TTCAACAAGTGGGGCAAGGGCCAGGCTGTACTACGACAGCATGAGCTAC GAGGACGAGGACGCGCTGTTGCGCTACACCAAGAGGAAAGCTGATCGCG ACCATCACCCAGATGAAGCTGAAGTACGCCATCAGCGCCAAAGAAAGAGG GCGAGGACCGCTGGCGCGCGCTGAGCATCTGCAGGACCATGACCAACAG CGAGTTCCACCAGAAGCTGCTGAAGAGCATGCGCGCCAGCAGGGCGCG CCACCGTGGTGATCGCGACCAAGCAAGTTCTACGCGCGCTGGCAACAAC TGGTGAAGACCGTGATACAGCGAGCTGGAGAACCCCACTGATGGGCT GGGACTACCCCAAGTGCGACAGGGCCATGCCCAACATGCTGAGGATCA TGGCCAGGCTGGTGCTGGCCAGGAAGGACACCACTGCTGCGAGCGTGA AGCCACAGGTTCTACAGGCTGGCCAAAGAGTGGCGCCAGGTGCTGAGC CGAGATGCTGATGTGCGCGCGGAGCTGTACGTGAAGCGCGCGCGGCA CGAGGACGCGCGCAGGCGCAGCAGCGCGCTACGCCAAGAGCGGTGTTCAAG ATCTGCCAGGCGCGTGACCGCCAAACGTGAACGCGCTGCTGAGGACCGA CGGCAACAAGATCGCGGACAAGTACCTGAGGAACCTGCAGGACAGGCT GTACGAGTGCGGTGTACAGGAACAGGGACCTGGACACCGACTTGGTGAA CGAGTCTACGCGTACCTGAGGAAGGAGCTCAGCATGATGATCTGAGG GAGGACGCGCGTGCTGTGCTTCAACAGCACCTAGGCCAGCGAGGGCGT GCTGGCCAGCATCAAGAACTTCAAGAGCGTGCTGTACTACCAAGAACAA GGTGTTGATGAGCGAGGCAAGTGTGACCGGAGACCGAGCGAGCTGACCA GGGCGCGCCAGGAGTTCTGCAAGCCAGGACACCATGCTGGTGAAAGCAGG GCGAGGACTACGTGTACCTGCCCCAGCCCGAGCCGAGGAGATGCTGG GCGCGCGCTGCTTCGTGGACGACATCGTGAAGACCGACCGGACCTGT ATGATCGAGAGGTTGCTGAGCCTGGCCATCGACGCGCTAGCCCGCTGAGC AAGCAGCCCAACCGAGGTACGCCAGCTGTGTCACCTGTGATGCTGAG TACATCAGGAAGCTGCACGACGAGCTGACCGGCGACATGCTGGACATG TACAGGCTGATGCTGACCAAGGACAAGACCGAGGAGTGTGGGAGCGCG GAGTCTACGAGGCGATGTACAGCGCGCACACCGCTGCTGAGCTGACTC <u>GAGCTGGTACTGCATGCACGCAATGCTAGCTGCCCCCTTCCCGTCTGTG</u> <u>GGTACCCCGAGTCTCCCCCGACCTCGGGTCCCAGGTATGCTCCCACC</u> <u>TCCACCTGCCCCACTCACCACCTCTGCTAGTTCCAGACACCTCCCAAG</u> <u>CACGCAGCAATGCAGCTCAAAACGCTTAGCCTAGCCACACCCCCACG</u> <u>GGAAACAGCAGTGATTAACCTTTAGCAATAAACGAAAGTTTAACTAAG</u> <u>CTATACTAACCCAGGGTTGGTCAATTTCTGTCAGCCACACCCCTGGA</u> <u>GCTAGCAAAAAAAAAA</u>	
9 <i>promoter, 5'UTR and leader sequence, Spike glycoprotein (HexaPro- mutations),</i>	CTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT AGTTCATAGCCCATATATGGAGTTCGCGGTACATAACTTACGGTAAATGG CCGCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAAT GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAAT GGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA TCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC GCCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCTACTTTGGCA GTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCA CGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCCATTTT GTATTTATTTATTTTAAATATTTGTGCAGCGATGGGGGCGGGGGGGG	147

linker. CXCL11, IL7. Stop Codon. 3'UTR and PolyA tail	GGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGG GGCGGGGCGAGGCGGAGAGGTGCGGCGGCGAGCCAATCAGAGCGGGCG CGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGGCGGGCGGGCCC TATAAAAAGCGAAGCGCGCGGGCGGGGAGTGGCTGCGCGCTGCCT TCGCCCCGTGCCCCGCTCCGCGCGCGGCTCGCGCGCGCGCGCGCGG CTCTGACTGACCGGCTTACTCCACAGGTGAGCGGGCGGGACGGCCC TTCTCCTCCGGGCTGTAAATAGCGCTTGGTTTAAATGACGGCTTGTTCCT TTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTG TGCGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTTGTGTGCGTGG GGAGCGCGCGTGC GGCTCCGCGCTGCCCGGGCGGCTGTGAGCGCTG CGGGCGCGGGCGGGGGCTTTGTGCGCTCCGCGAGTGTGCGCGAGGGG AGCGCGGGCGGGGGCGGTGCCCCGCGGTGCGGGGGGGGCTGCGAG GGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAG GGGGTGTGGGCGCGCTCGGTGCGGGCTGCAACCCCCCTGCAACCCCCCT CCCCGAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTAC GGGGCGTGGCGCGGGGCTCGCCGTGCGGGGCGGGGGGTGGCGGCA GGTGGGGGTGCGCGGGCGGGGCGGGGCGCGGCTCGGGCGGGGAGGG CTCGGGGAGGGGCGCGGGCGGGCGGGCGGGGAGCGCGGGCGGCTGTG AGGCGCGGGCGAGCGCGAGCCATTGCTTTTATGGTAATCGTGCGAGAG GGCGCAGGGACTTCCTTTGTCCCAAATCTGTGCGGAGCGGAAATCTGG GAGGCGCGCGCGCAACCCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGC GGCGCGGGCAGGAAGGAATGGGCGGGGAGGGGCTTCGTGCGTGCG CGCGCGCGGCTCCCCCTTCCTCCTCTCCAGCCTCGGGGGCTGTCGCGCG GGGACGGCTGCTTCGGGGGGGACGGGGCAGGGCGGGGTTCGGCT TCTGGCGTGTGACCGGGCGGCTCTAGAGCCTCTGCTAACCATTGTCATG CCTTCTTCTTTTCTTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCT GTCTCATCATTTTGGCAAAGAATTGGAGAATAAACTAGTATTCTTCTGGT CCACAGACTCAGAGAGAACC CGCCACCATGTTCTGTGTTCTGGTGCT GCTGCCCCCTGGTGAGCAGCCAGTGC GTGAACCTGACCACAGGACCC AGCTGCCCCCGCTACACCAACAGCTTCACCAGGGGCGGTGTACTACC CCGACAAGGTGTT CAGGAGCAGCGTGTGCACAGCACCCAGGACCTG TTCCTGCCCTTCTCAGCAACGTGACCTGGTTCCACGCCATCCACGTGA GCGGCACCAACGGCACCAAGAGGTTGACAAACCCCGTGTGCTGCCCTT AACGACGGCGTGACTTCGCCAGCACCGAGAAGAGCAACATCATCAGG GGCTGGATCTTCGGCACCAACCTGGACAGCAAGACCCAGAGCCTGCT GATCGTGAACAACGCCACCAACGTGGTGATCAAGGTGTGCGAGTTCCA GTTCTGCAACGACCCCTTCCTGGGCGTGTACTACCACAAGAACAACAA GAGCTGGATGGAGAGCGAGTT CAGGGTGACAGCAGCGCCAACT GCACCTTCGAGTACGTGAGCCAGCCCTTCCTGATGGACCTGGAGGGCA AGCAGGGCAACTTCAAGAACCTGAGGGAGTTCTGTTCAAGAACATCG ACGGCTACTTCAAGATCTACAGCAAGCACACCCCATCAACCTGGTGAG GGACCTGCCCCAGGGCTTCAGCGCCCTGGAGCCCTGGTGACCTGC CCATCGGCATCAACATCACCAGGTTCCAGACCCTGCTGGCCCTGCACA GGAGCTACCTGACCCCGGGCAGAGCAGCAGCGGCTGGACCGCCGG CGCCGCCCGCTACTACGTGGGCTACCTGCAGCCAGGACCTTCCTGCT GAAGTACAACGAGAACGGCACCATCACCGACGCCGTGGACTGCGCCCT GGACCCCTGAGCGAGACCAAGTGACCCCTGAAGAGCTTCACCGTGG AGAAGGGCATCTACCAGACCAGCAACTTCAGGGTGACGCCACCGAGA GCATCGTGAGGTTCCCAACATCACCACCTGTGCCCCCTTCGGCGAGG TGTTCAACGCCACCAAGTTCCGCCAGCGGTGACGCCCTGGAACAGGAAGA GGATCAGCAACTGCGTGGCCGACTACAGCGTGTGTACAACAGCGCCA GCTTCAGCACCTTCAAGTGCTACGGCGTGAGCCCCACCAAGCTGAACG ACCTGTGCTTCACCAACGTGTACGCCGACAGCTTCGTGATCAGGGGCG ACGAGGTGAGGCAGATCGCCCCCGGCCAGACCGGCAAGATCGCCGAC TAACTACAAGCTGCCCCGACGACTTCACCGGCTGCGTGATCGCCTGG AACAGCAACAACCTGGACAGCAAGGTGGGCGGCACTACAACCTG TACAGGCTGTT CAGGAAGAGCAACCTGAAGCCCTTCGAGAGGGACATC AGCACCGAGATCTACCAGGCCGGCAGCACCCCTGCAACGGCGTGGA
--	--

	GGGCTTCAACTGCTACTTCCCCCTGCAGAGCTACGGCTTCCAGCCCAC CAACGGCGTGGGCTACCAGCCCTACAGGGTGGTGGTGTGAGCTTCG AGCTGCTGCACGCCCCCGCCACCGTGTGCGGCCCCAAGAAGAGCACC AACCTGGTGAAGAACAAGTGCGTGAACCTCAACTTCAACGGCCTGACC GGCACCGGCGTGCTGACCGAGAGCAACAAGAAGTTCCTGCCCTTCCA GCAGTTCGGCAGGGACATCGCCGACACCACCGACGCCGTGAGGGACC CCCAGACCCTGGAGATCCTGGACATCACCCCCCTGCAGCTTCGGCGGC GTGAGCGTGATACCCCCCGGCACCAACACCAGCAACCAGGTGGCCGT GCTGTACCAGGACGTGAACTGCACCGAGGTGCCCGTGGCCATCCACG CCGACCAGCTGACCCCCACCTGGAGGGTGTACAGCACCGGCAGCAAC GTGTTCCAGACCAGGGCCGGCTGCCTGATCGGCGCCGAGCACGTGAA CAACAGCTACGAGTGCGACATCCCCATCGGCGCCGGCATCTGCGCCAG CTACCAGACCCAGACCAACAGCCCCGGCAGCGCGCCAGCGTGGCCA GCCAGAGCATCATCGCCTACACCATGAGCCTGGGCGCCGAGAACAGCG TGGCCTACAGCAACAACAGCATCGCCATCCCCACCAACTTCAACATCAG CGTGACCACCGAGATCCTGCCCGTGAGCATGACCAAGACCAGCGTGGA CTGCACCATGTACATCTGCGGCGACAGCACCGAGTGCAGCAACCTGCT GCTGCAGTACGGCAGCTTCTGCACCCAGCTGAACAGGGCCCTGACCG GCATCGCCGTGGAGCAGGACAAGAACACCCAGGAGGTGTTCCGCCAG GTGAAGCAGATCTACAAGACCCCCCCCCATCAAGGACTTCGGCGGGCTTC AACTTCAGCCAGATCCTGCCCGACCCCCAGCAAGCCAGCAAGAGGAG CCCCATCGAGGACCTGCTGTTCAACAAGGTGACCCTGGCCGACGCCG GCTTCATCAAGCAGTACGGCGACTGCCTGGGCGACATCGCCGCCAGG GACCTGATCTGCGCCCAGAAGTTCAACGGCCTGACCGTGCTGCCCCC CTGCTGACCGACGAGATGATCGCCAGTACACCAGCGCCCTGCTGGCC GGCACCATCACCAGCGGCTGGACCTTCGGCGCCGGCCCGCCGCTGCA GATCCCCCTTCCCATGCAGATGGCCTACAGGTTCAACGGCATCGGCGT GACCCAGAACGTGCTGTACGAGAACCAGAAAGCTGATCGCCAACCAAGT CAACAGCGCCATCGGCAAGATCCAGGACAGCCTGAGCAGCACCCCGCA GCGCCCTGGGCAAGCTGCAGGACGTGGTGAACCAGAACGCCAGGCC CTGAACACCCTGGTGAAGCAGCTGAGCAGCAACTTCGGCGCCATCAGC AGCGTGCTGAACGACATCCTGAGCAGGCTGGACCCCCGAGGCCGA GGTGACATCGACAGGCTGATCACCGGCAGGCTGCAGAGCCTGCAGA CCTACGTGACCCAGCAGCTGATCAGGGCCGCCGAGATCAGGGCCAGC GCCAACCTGGCCGCCACCAAGATGAGCGAGTGCGTGCTGGGCCAGAG CAAGAGGGTGGACTTCTGCGGCAAGGGCTACCACCTGATGAGCTTCCC CCAGAGCGCCCCCACGGCGTGGTGTTCCTGCAGTGACCTACGTGTC CCGCCCAGGAAAGAACTTCACCACCGCCCCCGCCATCTGCCACGAC GGCAAGGCCCACTTCCCCAGGGAGGGCGTGTTCGTGAGCAACGGCAC CCTGTTGCTGACCCAGAGGAACTTCTACGAGCCCCAGATCATCAC CACCAGAACACCTTCGTGAGCGGCAACTGCGACGTGGTGATCGGCAT CGTGAACAACACCGTGACGACCCCTGCAGCCCCAGCTGGACAGCTT CAAGGAGGAGCTGGACAAGTACTTCAAGAACCACACCAGCCCCGACGT GGACCTGGGCGACATCAGCGGCATCAACGCCAGCGTGGTGAACATCCA GAAGGAGATCGACAGGCTGAACGAGGTGGCCAAGAACCTGAACGAGA GCCTGATCGACCTGCAGGAGCTGGGCAAGTACGAGCAGTACATCAAGT GGCCCTGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCATCG TGATGGTGACCATCATGCTGTGCTGCATGACCAGCTGCTGCAGCTGCCT GAAGGGCTGCTGCAGCTGCGGCAGCTGCTGCAAGTTCGACGAGGACG ACAGCGAGCCCGTGCTGAAGGGCGTGAAGCTGCACTACACCGGAAGC GGAGCCACGAACCTTCTCTCTGTTAAAGCAAGCAGGAGATGTTGAAGAAA <u>ACCCCGGGCCATGAACAGGAAGGTGACCGCCATCGCCCTGGCCGCC</u> ATCATCTGGGCCACCGCCGCCAGGGCTTCTGATGTTCAAGCAGGG CAGGTGCCTGTGCATCGGCCCCGGCATGAAGGCCGTGAAGATGGCCG AGATCGAGAAGGCCAGCGTGATCTACCCAGCAACGGCTGCGACAAG GTGAGGTGATCGTGACCATGAAGGCCACAAGAGGCAGAGGTGCC TGGACCCAGGAGCAAGCAGGCCAGGCTGATCATGCAGGCCATCGA GAAGAAGAACTTCTGAGGAGGCAGAACATGTGAGGAAGCGGAGCC	
--	---	--

	<u>ACGAACCTTCTCTCTGTTAAAGCAAGCAGGAGATGTTGAAGAAAACCCCG</u> <u>GGCCTATGTTCCACGTGAGCTTCAGGTACATCTTCGGCATCCCCCCC</u> <u>TGATCCTGGTGCTGCTGCCCCGTGACCAGCAGCGAGTGCCACATCAAG</u> <u>GACAAGGAGGGCAAGGCCTACGAGAGCGTGCTGATGATCAGCATCGA</u> <u>CGAGCTGGACAAGATGACCGGCACCGACAGCAACTGCCCCAAACAAC</u> <u>GAGCCCAACTTCTTCAGGAAGCACGTGTGCGACGACACCAAGGAGG</u> <u>CCGCCTTCCTGAACAGGGCCGCCAGGAAGCTGAAGCAGTTCCTGAAG</u> <u>ATGAACATCAGCGAGGAGTTCAACGTGCACCTGCTGACCGTGAGCCA</u> <u>GGGCACCCAGACCTGCTGAACTGCACCAGCAAGGAGGAGAAGAAC</u> <u>GTGAAGGAGCAGAAGAAGAACGACGCCTGCTTCCTGAAGAGGCTGC</u> <u>TGAGGGAGATCAAGACCTGCTGGAACAAGATCCTGAAGGGCAGCATC</u> <u>TGATGACTCGAGCTGGTACTGCATGCACGCAATGCTAGCTGCCCTTT</u> <u>CCCGTCCTGGGTACCCCGAGTCTCCCCGACCTCGGGTCCCAGGTAT</u> <u>GCTCCCACTCCACCTGCCCACTCACCACCTCTGCTAGTTCAGACA</u> <u>CCTCCCAAGCACGCAGCAATGCAGCTCAAAACGCTTAGCCTAGCCAC</u> <u>ACCCCCACGGGAAACAGCAGTGATTAACCTTTAGCAATAAACGAAAGT</u> <u>TTAACTAAGCTATACTAACCCAGGGTTGGTCAATTTTCGTGCCAGCCAC</u> <u>ACCTTGGAGCTAGCAAAAAAAAA</u>	
--	--	--

[00276] As mentioned above, the present invention is not limited to the examples in Table 9. In some embodiments, vaccine candidates may comprise various pieces (e.g. promoters, proteins, adjuvants) as shown described herein.

[00277] Table 10 shows non-limited examples of proteins that may be used to create a vaccine composition described herein. In some embodiments, proteins listed below may be arranged in a plurality of combinations. In some embodiments, the proteins may be directly linked together. In other embodiments, the proteins are linked together via a linker.

[00278] Table 10 shows non-limiting examples of spike proteins.

Table 10

Proteins	Sequence	SEQ ID NO:
Spike glycoprotein with 6 stabilizing mutations (HexaPro)	ATGTTTCGTGTTTCTGGTGCTGCTGCCCCCTGGTGAGCAGCCAGTG CGTGAACCTGACCACCCGGACCCAGCTGCCACCAGCCTACACC AACAGCTTCACCCGGGGCGTCTACTACCCGACAAGGTGTTCCG GAGCAGCGTCTGCACAGCACCCAGGACCTGTTCTGCCCCTTCT TCAGCAACGTGACCTGGTTCCACGCCATCCACGTGAGCGGCACC AACGGCACCAAGCGGTTTCGACAACCCCGTGCTGCCCTTCAACGA CGGCGTGTAATTCGCCAGCACCCGAGAAGAGCAACATCATCCGGG GCTGGATCTTCGGCACCACCCTGGACAGCAAGACCCAGAGCCT GCTGATCGTGAATAACGCCACCAACGTGGTGATCAAGGTGTGCG AGTTCCAGTTCTGCAACGACCCCTTCTGGGCGTGTAATACAC AAGAACAACAAGAGCTGGATGGAGAGCGAGTTCGGGTGTACAG CAGCGCCAACAACCTGCACCTTCAGTACGTGAGCCAGCCCTTCC TGATGGACCTGGAGGGCAAGCAGGGCAACTTCAAGAACCTGCG GGAGTTCGTGTTCAAGAACATCGACGGCTACTTCAAGATCTACAG CAAGCACACCCCAATCAACCTGGTGCGGGATCTGCCCCAGGGCT TCTCAGCCCTGGAGCCCTGGTGACCTGCCATCGGCATCAAC ATCACCCGGTTCCAGACCCTGCTGGCCCTGCACCGGAGCTACCT GACCCAGGCGACAGCAGCAGCGGGTGGACAGCAGGCGCGGC	148

	TGCTTACTACGTGGGCTACCTGCAGCCCCGGACCTTCCTGCTGA AGTACAACGAGAACGGCACCATCACCGACGCCGTGGACTGCGC CCTGGACCCTCTGAGCGAGACCAAGTGCACCCTGAAGAGCTTCA CCGTGGAGAAGGGCATCTACCAGACCAGCAACTTCCGGGTGCA GCCCCACGAGAGCATCGTGCGGTTCCCCAACATCACCAACCTGT GCCCCCTTCGGCGAGGTGTTCAACGCCACCCGGTTCGCCAGCGT GTACGCCTGGAACCGGAAGCGGATCAGCAACTGCGTGGCCGAC TACAGCGTGCTGTACAACAGCGCCAGCTTCAGCACCTTCAAGTG CTACGGCGTGAGCCCCACCAAGCTGAACGACCTGTGCTTCACCA ACGTGTACGCCGACAGCTTCGTGATCCGTGGCGACGAGGTGCG GCAGATCGCACCCGGCCAGACAGGCAAGATCGCCGACTACAAT ACAAGCTGCCCCGACGACTTCACCGGCTGCGTGATCGCCTGGAA CAGCAACAACCTCGACAGCAAGGTGGCGGGCAACTACAATACC TGTAACGGCTGTTCCGGAAGAGCAACCTGAAGCCCTTCGAGCG GGACATCAGCACCGAGATCTACCAAGCCGGCTCCACCCCTTGCA ACGGCGTGGAGGGCTTCAACTGCTACTTCCCTCTGCAGAGCTAC GGCTTCAGCCCCACCAACGGCGTGGGCTACCAGCCCTACCGGG TGGTGGTGCTGAGCTTCGAGCTGCTGCACGCCCCAGCCACCGT GTGTGGCCCCAAGAAGAGCACCAACCTGGTGAAGAACAAGTGC GTGAACCTCAACTTCAACGGCCTTACCGGCACCGGCGTGCTGAC CGAGAGCAACAAGAAATTCTGCCCTTTCAGCAGTTTCGGCCGGG ACATCGCCGACACCACCGACGCTGTGCGGGATCCCCAGACCCT GGAGATCCTGGACATCACCCCTTGACGCTTCGGCGGCGTGAGC GTGATCACCCAGGCACCAACACCAGCAACCAGGTGGCCGTGC TGTACCAGGACGTGAAGTGCACCGAGGTGCCCGTGGCCATCCA CGCCGACCAGCTGACACCCACCTGGCGGGTCTACAGCACCGGC AGCAACGTGTTCCAGACCCGGGCGCGTTCCTGATCGGCGCGG AGCACGTGAACAACAGCTACGAGTGCGACATCCCCATCGGCGCC GGCATCTGTGCCAGCTACCAGACCCAGACCAATTCACCCGGCAG CGCCAGCAGCGTGGCCAGCCAGAGCATCATCGCTACACCATGA GCCTGGGCGCCGAGAACAGCGTGGCCTACAGCAACAACAGCAT CGCCATCCCCACCAACTTCACCATCAGCGTGACCAACGAGATT TGCCCGTGAGCATGACCAAGACCAGCGTGGACTGCACCATGTAC ATCTGCGGCGACAGCACCGAGTGACGCAACCTGCTGCTGCAGTA CGGCAGCTTCTGCACCCAGCTGAACCGGGCCCTGACCGGCATC GCCGTGGAGCAGGACAAGAACACCCAGGAGGTGTTGCCCCAGG TGAAGCAGATCTACAAGACCCCTCCCATCAAGGACTTCGGCGGC TTCAACTTCAGCCAGATCCTGCCCGACCCAGCAAGCCAGCAA GCGGAGCCCATCGAGGACCTGCTGTTCAACAAGGTGACCCTAG CCGACGCCGGCTTCATCAAGCAGTACGGCGACTGCCTCGGCGA CATAGCCGCCCCGGGACCTGATCTGCGCCAGAAAGTTCAACGGCC TGACCGTGCTGCCTCCCCTGCTGACCGACGAGATGATCGCCAG TACACCAGCGCCCTGTTAGCCGGAACCATCACAGCGGCTGGAC TTTCGGCGCTGGCCGCTCTGCAGATCCCCTTCCCATGCAGA TGGCCTACCGGTTCAACGGCATCGGCGTGACCCAGAACGTGCT GTACGAGAACCAGAAGCTGATCGCCAACCAAGTTCAACAGCGCCA TCGGCAAGATCCAGGACAGCCTGAGCAGCACCCATAGCGCCCT GGGCAAGCTGCAGGACGTGGTGAACCAGAACGCCAGGCCCTG AACACCCTGGTGAAGCAGCTGAGCAGCAACTTCGGCGCCATCAG CAGCGTGCTGAACGACATCCTGAGCCGGCTGGAACCTCCGAG GCCGAGGTGCAGATCGACCGGCTGATCACTGGCCGGCTGCAGA GCCTGCAGACCTACGTGACCCAGCAGCTGATCCGGGCGGCCGA GATTCGGGCCAGCGCCAACCTGGCCGCCACCAAGATGAGCGAG TGCTGTGCTGGGCCAGAGCAAGCGGGTGGACTTCTGCGGCAAGG GCTACCACCTGATGAGCTTCCCCAGAGCGCACCCACGGAGTG GTGTTCTGACGTGACCTACGTGCCCGCCAGGAGAAGAATT CACCACCGCCCGACCATCTGCCACGACGGCAAGGCCACTTT CCCCGGGAGGGCGTGTTCTGAGCAACGGCACCCACTGTTTCG	
--	---	--

	TGACCCAGCGGAACTTCTACGAGCCCCAGATCATCACCACCGAC AACACCTTCGTGAGCGGCAACTGCGACGTGGTGATCGGCATCGT GAACAACACCGTGTACGATCCCCCTGCAGCCCCGAGCTGGACAGCT TCAAGGAGGAGCTGGACAAGTACTTCAAGAATCACACCAGCCCC GACGTGGACCTGGGCGACATCAGCGGCATCAACGCCAGCGTGG TGAACATCCAGAAGGAGATCGATCGGCTGAACGAGGTGGCCAAAG AACCTGAACGAGAGCCTGATCGACCTGCAGGAGCTGGGCAAGTA CGAGCAGTACATCAAGTGGCCCTGGTACATCTGGCTGGGCTTCA TCGCCGGCCTGATCGCCATCGTGATGGTGACCATCATGCTGTGC TGCATGACCAGCTGCTGCAGCTGCCTGAAGGGCTGTTGCAGCTG CGGCAGCTGCTGCAAGTTCGACGAGGACGACAGCGAGCCCCGTG CTGAAGGGCGTGAAGCTGCACTACACCTGATAATAGGCTGGAGC CTCGGTGGCCTAGCTTCTTGGCCCTTGGGCTCCCCCAGCCCC TCCTCCCCTTCTGCACCCGTACCCCGTGGTCTTTGAATAAAGT CTGAGTGGGCGGCAAAAAAAAAA	
Spike glycoprotein with one stabilizing mutations	ATGTTCTGTGTTCTGGTGCTGCTGCCCCCTGGTGAGCAGCCAGTG CGTGAACCTGACCACCCGGACCCAGCTGCCACCAGCCTACACC AACAGCTTCACCCGGGGCGTCTACTACCCCGACAAGGTGTTCCG GAGCAGCGTCTGACAGCAGCAGGACCTGTTCTGCCCTTCT TCAGCAACGTGACCTGGTTCCACGCCATCCACGTGAGCGGCACC AACGGCACCAAGCGTTTCGACAACCCCGTGCTGCCCTTCAACGA CGGCGTGACTTCGCCAGCAGCAGAGAAGAGCAACATCATCCGGG GCTGGATCTTCGGCACCACCCTGGACAGCAAGACCCAGAGCCT GCTGATCGTGAATAACGCCACCAACGTGGTGATCAAGGTGTGCG AGTTCCAGTTCTGCAACGACCCCTTCTGGGCGTGTAACAC AAGAACAACAAGAGCTGGATGGAGAGCGAGTTCGGGTGTACAG CAGCGCCAACAACCTGCACCTTCGAGTACGTGAGCCAGCCCTTCC TGATGGACCTGGAGGGCAAGCAGGGCAACTTCAAGAACCTGCG GGAGTTCGTGTTCAAGAACATCGACGGCTACTTCAAGATCTACAG CAAGCACACCCCAATCAACCTGGTGCGGGATCGCCCCAGGGCT TCTCAGCCCTGGAGCCCCTGGTGGACCTGCCATCGGCATCAAC ATCACCCGGTTCCAGACCCTGCTGGCCCTGCACCGAGCTACCT GACCCAGGCGACAGCAGCAGCGGTGGACAGCAGGCGCGGC TGCTTACTACGTGGGTACCTGCAGCCCCGACCTTCTGCTGA AGTACAACGAGAAGCGCACCATCACCGACGCCGTGGACTGCGC CCTGGACCCTCTGAGCGAGACCAAGTGCACCCTGAAGAGTTCA CCGTGGAGAAAGGGCATCTACCAGACCAGCAACTTCCGGGTGCA GCCCACCGAGAGCATCGTGCGGTTCCTCAACATCAACCACTGT GCCCCCTTCGGCGAGGTGTTCAACGCCACCCGTTTCGCCAGCGT GTACGCCTGGAACCGGAAGCGGATCAGCAACTGCGTGGCCGAC TACAGCGTGCTGTACAACAGCGCCAGCTTACGACCTTCAAGTG CTACGGCGTGAGCCCCACCAAGCTGAACGACCTGTGCTTACCA ACGTGTACGCCGACAGCTTCGTGATCCGTGGCGACGAGGTGCG GCAGATCGCACCCGGCCAGACAGGCAAGATCGCCGACTACAAC ACAAGCTGCCCCGACGACTTCACCGGCTGCGTGATCGCCTGGAA CAGCAACAACCTCGACAGCAAGGTGGGCGGCAACTACAAC TGTAACCGGCTGTTCCGGAAGAGCAACCTGAAGCCCTTCGAGCG GGACATCAGCACCGAGATCTACCAAGCCGGCTCCACCCCTTGCA ACGGCGTGGAGGGCTTCAACTGCTACTTCCCTCTGCAGAGCTAC GGCTTCCAGCCCACCAACGGCGTGGGCTACCAGCCCTACCGGG TGGTGGTGCTGAGCTTCGAGCTGCTGCACGCCCCAGCCACCGT GTGTGGCCCCAAGAAGAGCACCAACCTGGTGAAGAACAAGTGC GTGAACCTCAACTTCAACGGCCTTACCGGCACCGGCGTGCTGAC CGAGAGCAACAAGAAATTCCTGCCCTTTCAGCAGTTGGGCCGGG ACATCGCCGACACCGACGCTGTGCGGGATCCCCAGACCT GGAGATCCTGGACATACCCCTTGCAGCTTCGGCGGCGTGAGC GTGATCACCCAGGCACCAACACCAGCAACCAGGTGGCCGTGC	149

	TGTACCAGGACGTGAACTGCACCGAGGTGCCCCGTGGCCATCCA CGCCGACCAGCTGACACCCACCTGGCGGGTCTACAGCACCGGC AGCAACGTGTTCCAGACCCGGGCGCGTTGCCTGATCGGCGCCG AGCACGTGAACAACAGCTACGAGTGCGACATCCCCATCGGCGCC GGCATCTGTGCCAGCTACCAGACCCAGACCAATTCACCCGGCAG CGCCAGCAGCGTGGCCAGCCAGAGCATCATCGCTACACCATGA GCCTGGGCGCCGAGAACAGCGTGGCCTACAGCAACAACAGCAT CGCCATCCCCACCAACTTCACCATCAGCGTGACCACCGAGATT TGCCCGTGAGCATGACCAAGACCAGCGTGGACTGCACCATGTAC ATCTGCGGCGACAGCACCGAGTGACAGCAACCTGCTGCTGCAGTA CGGCAGCTTCTGCACCCAGCTGAACCGGGCCCTGACCGGCATC GCCGTGGAGCAGGACAAGAACACCCAGGAGGTGTTGCGCCAGG TGAAGCAGATCTACAAGACCCCTCCCATCAAGGACTTCGGCGGC TTCAACTTCAGCCAGATCCTGCCCCGACCCAGCAAGCCAGCAA GCGGAGCCCCATCGAGGACCTGCTGTTCAACAAGGTGACCCTAG CCGACGCCGGCTTCATCAAGCAGTACGGCGACTGCCTCGGCGA CATAGCCGCCCCGGGACCTGATCTGCGCCAGAAAGTTCAACGGCC TGACCGTGCTGCCTCCCCTGCTGACCGACGAGATGATCGCCAG TACACCAAGCGCCCTGTTAGCCGGAACCATCACAGCGGCTGGAC TTTCGGCGCTGGCCCCCGCTCTGCAGATCCCCTTCCCCATGCAGA TGCCCTACCGGTTCAACGGCATCGGCGTGACCCAGAACGTGCT GTACGAGAACCAGAAAGCTGATCGCCAAACAGTTCAACAGCGCCA TCGGCAAGATCCAGGACAGCCTGAGCAGCACCCCTAGCGCCCT GGGCAAGCTGCAGGACGTGGTGAACCAGAACGCCCCAGGCCCTG AACACCCTGGTGAAGCAGCTGAGCAGCAACTTCGGCGCCATCAG CAGCGTGCTGAACGACATCCTGAGCCGGCTGGACCCCCGAG GCCGAGGTGCAGATCGACCGGCTGATCACTGGCCGGCTGCAGA GCCTGCAGACCTACGTGACCCAGCAGCTGATCCGGGCGCCGA GATTGCGGCCAGCGCCAACTGGCCGCCACCAAGATGAGCGAG TGCCTGCTGGGCCAGAGCAAGCGGGTGGACTTCTGCGGCAAGG GCTACCACCTGATGAGCTTTCCCCAGAGCGCACCCACCGGAGTG GTGTTCTGTCAGTGACCTACGTGCCCGCCAGGAGAAGAATT CACCAACGCCCCAGCCATCTGCCACGACGGCAAGGCCCACTTT CCCCGGGAGGGCGTGTTCTGAGCAACGGCACCCACTGGTTCTG TGACCCAGCGGAATTCTACGAGCCCCAGATCATCACCAACGAC AACACCTTCGTGAGCGGCAACTGCGACGTGGTGATCGGCATCGT GAACAACACCGTGATGATCCCCTGCAGCCCCAGCTGGACAGCT TCAAGGAGGAGCTGGACAAGTACTTCAAGAAATCACACCAACCCC GACGTGGACCTGGGCGACATCAGCGGCATCAACGCCAGCGTGG TGAACATCCAGAAGGAGATCGATCGGCTGAACGAGGTGGCCAG AACCTGAACGAGAGCCTGATCGACCTGCAGGAGCTGGGCAAGTA CGAGCAGTACATCAAGTGGCCCTGGTACATCTGGCTGGGCTTCA TCGCCGGCCTGATCGCCATCGTGATGGTGACCATCATGCTGTGC TGCATGACCAGCTGCTGCAGCTGCCTGAAGGGCTGTTGCAGCTG CGGCAGCTGCTGCAAGTTCGACGAGGACGACAGCGAGCCCCGTG CTGAAGGGCGTGAAGCTGCACTACACCTGATAATAGGCTGGAGC CTCGGTGGCCTAGCTTCTTGCCCCCTGGGCCCTCCCCCAGCCCC TCCTCCCCTTCTGCACCCGTACCCCCGTGGTCTTTGAATAAAGT CTGAGTGGGCGGCAAAAAAAAAA	
Nucleocapsid	ATGAGCGACAACGGCCCCCAGAACCAGAGGAACGCCCCAGGA TCACCTTCGGCGGGCCCCAGCGACAGCACCGGCAGCAACCAGAA CGGCGAGAGGAGCGGCGCCAGGAGCAAGCAGAGGAGGCCCA GGGCTGCCCAACAACACCGCCAGCTGGTTACCGCCCTGACC CAGCACGGCAAGGAGGACCTGAAGTTCCCCAGGGGCCAGGGC GTGCCCATCAACACCAACAGCAGCCCCGACGACCATGCTGTA CTACAGGAGGGCCACCAGGAGATCAGGGGCGGCGACGGCAA GATGAAGGACCTGAGCCCCAGGTGGTACTTCTACTACCTGGCA	150

	CCGGCCCCGAGGCCGGCCTGCCCTACGGCGCCAACAAGGACG GCATCATCTGGGTGGCCACCGAGGGCGCCCTGAACACCCCCAA GGACCACATCGGCACCAAGGAACCCCGCCAACAACGCCGCCATC GTGCTGCAGCTGCCCCAGGGCACCACCCTGCCCCAAGGCTTCT ACGCCGAGGGCAGCAGGGGGCGGCAGCCAGGCCAGCAGCAGGA GCAGCAGCAGGAGCAGGAACAGCAGCAGGAACAGCACCCCCG GCAGCAGCAGGGGCACCAGCCCCGCCAGGATGGCCGGCAACG GCGGCGACGCCGCCCTGGCCCTGCTGCTGCTGGACAGGCTGAA CCAGCTGGAGAGCAAGATGAGCGGCAAGGGCCAGCAGCAGCAG GGCCAGACCGTGACCAAGAAGAGCGCCGCCGAGGCCAGCAAG AAGCCCAGGCAGAAGAGGACCGCCACCAAGGCCTACAACGTGA CCCAGGCCCTTCGGCAGGAGGGGGCCCCGAGCAGACCCAGGGCA ACTTCGGCGACCAGGAGCTGATCAGGCAGGGCACCGACTACAA GCACTGGCCCCAGATCGCCCAAGTTGCCCCCAGCGCCAGCGCC TTCTTCGGCATGAGCAGGATCGGCATGGAGGTGACCCCCAGCG GCACCTGGCTGACCTACACCGGCGCCATCAAGCTGGACGACAA GGACCCCAACTTCAAGGACCAGGTGATCCTGCTGAACAAGCACA TCGACGCCTACAAGACCTTCCCCCCCACCGAGCCCAAGAAGGAC AAGAAGAAGAAGGCCGACGAGACCCAGGCCCTGCCCCAGAGGC AGAAGAAGCAGCAGACCGTGACCCTGCTGCCCCGCCGCCACCT GGACGACTTCAGCAAGCAGCTGCAGCAGAGCATGAGCAGCGCC GACAGCACCCAGGCC	
ORF1ab (non-annotated)	CAAACCACTGAAACAGCWCACTCTTGTAATGTTAACCGCTTTAAT GTGGCTATTACAAGAGCAAAAATTGGCATTGTTGTCATAATGTCTG ACAGAGATCTTTATGACAAGCTGCAATTCACAAGTCTAGAAGTACC GCGTCGTAACGTGGCTACATTACAAGCGGAAAATGTAAGTGGACT CTTTAAGGACTGTAGTAAGATCATAACTGGTCTTCATCCTACACAA GCACCTACACACCTTAGTGTTGATACAAAATTCAAGACTGAGGGA CTATGTGTTGACATACCAGGCATWCCWAAGGACATGACCTATMG WAGACTCATCTCYATGATGGGTTTCAAAATGAATTAYCAAGTTAAT GGTTACCCTAAYATGTTYATCACCCGYGARGAAGCCATMMGMCA GTWCGTGATGGATTGGCTTTGATGTAGAGGGKGTGTCATGCTACT AGGGATGCTGTCGGTACTAACCTACCTCTCCAGTTAGGATTTTCTA CAGGTGTTAACTTAGTAGCTGTACCAACTGGCTATGTTGACACTG AAAAAATACAGAATTCACCAGAGTTAATGCAAAACCTCCACCAG GTGACCAATTTAAACATCTTATACCACTTATGTACAAAGGTTTACCC TGGAACATAGTGCATCAAGATAGTACAAATGCTCAGTGATACAC TGAAAGGATTATCRGACAGAGTTGTGTTTGTCTATGGGCACATG GCTTTGAACCTACATCAATGAAGTACTTTGTCAAGATTGGACCTGA AAGAACGTGTTGTCTGTGTGACAAACGTGCAACTTGTTTTTCTAC TTCATCAGACAATTATGCCTGCTGGAACCATTTCTGTGGGTTTTGA CTATGTCTATAATCCATTTATGATTGATGTCCAGCAGTGGGGTTTTA CAGGTAACCTTCAGAGTAATCAGGATCAGCATTGCCAAGTGCATG GCAACGCTCATGTGGCTAGTTGTGATGCTATCATGACTAGATGTTT AGCAGTCCATGAGTGCTTTGTTAAGCGCGTTGACTGGTCTGTTGA GTACCCAATTATAGGTGATGAACCTGAAGATCAATGCCGATGCAG AAAAGTGCAACATATGGTTGTAAAGTCTGCATTGCTTGCTGACAA ATTCCCAGTTCTTCATGACATTGGAAACCCAAAGGCTATCAAATGT GTCCCRCAAGGCTGAAGTGGAATTGGAAGTTCTATGATGCTCAGCC CTGCAGTGACAAAGCTTATAAAATAAAAGAACTCTTCTATTCTTATG CTACACATCATGATAAATTCATTGATGGTGTTTGTATTTTGGAAAT TGTAACGTTGATCGTTACCCTGCCAATGCTATTGTRTGCAGGTTT GACACGAGAGTCTTGTCAAATTTGAACTTGCCAGGTTGTGATGGT GGTAGTTTGATGTAAATAAGCATGCATTCCACACTCCAGCTTTTG ATAAAAGTGCACTTACTAATTTAAAGCAATTGCCTTTCTTTTACT CTGACAGTCCCTGTGAGTCACATGGCAAGCAGGTTGTTTCTGAC ATTGATTATGTACCACTCAAATCTGCTACRTGTATAACACGATGCAA	151

	TTTGGGRGGTGCTGTTTGCAGACATCATGCAAATGAGTACCGACA GTACTTGGATGCATACAATATGATGATTTCTGCTGGCTTTAGCCTC TGGATTTACAAACAGTTTGACACTTATAACCTGTGGAACACCTTTA CCAGGTTACAGAGTTTAGAAAATGTGGCTTACAATGTTGTTAACAA AGGACACTTCGATGGACAAGCTGGTGAAGCACCTGTTCCGTCA TTAATAATGTTGTTTACACAAAGGTAGATGGTGTGATGTAGAGAT CTTTGAAAACAAGACAACACTTCCTGTTAATGTTGCATTTGAGCTT TGGGCTAAGCGTAACATTAAACCAGTGCCAGAGATTAAGATACTC AATAATTTGGGTGTCGATATCGCTGCTAATACTGTAATCTGGGACT ACAAGAGAGAAGCACCAGCACATATGTCAACAATAGGTGTCTGCA CAATGACTGACATTGCCAAGAAACCTACTGAGAGTGCTTGTTCTCCT CGCTTACTGTCTTATTTGATGGTAGAGTGGGAAGGACAGGTAGACC TTTTTAGAAATGCCCGTAATGGTGTGTTTAAATAACAGAAGGTTTCA TAAAGGTTTAAATACCTTCAAAGGGACCAGCACAAAGCTAGTGTCAA TGGAGTCACATTAATTGGAGAATCAGTAAAAACACAGTTTAAATTAT TTAAGAAAGTAGATGGCATCATTCAACAGTTGCCTGAAACCTACTT TACTCAGAGCCGAGACTTAGAGGATTTCAAGCCAGATCACAAAT GAAACTGACTTTCTTGAGCTCGCTATGGATGAATTCATACAACG GTACAAGCTTGAAGGCTATGCCTTCGAACATATCGTTTATGGAGAT TTTAGTCATGGACAGCTTGGTGGACTTCATCTAATGATTGGTCTAG CTAAGCGCTCACAAAGATTCACCACTTAAATTAGAGGATTTTATCCC TACGGACAGTACAGTAAAAATTATTTTATAACAGATGCGCAAAACA GGTTCATCAAAATGCGTGTGCTCTGTTATTGATCTTCTGCTTGATG ACTTTGTTGAGATAATAAAGTCACAAGATTTATCAGTGGTTTCAA GGTGGTCAAAGTCACAATTGACTATGCTGAAATTTTATTTCATGTTA TGGTGTAAGGATGGACATGTTGAAACCTTTTACCACAAATACAAAG CGAGTCAGGCGTGGCAACCAGGAGTTGCAATGCCTAACTTGTAT AAGATGCAGAGAATGCTTCTTGAAAAATGTGACCTTCAGAATTATG GTGAAAAATGCTGTCATACCAAARGGAATAATGATGAATGTCGCAAA ATATACTCAACTGTGTCAATATTTAAATACACTYACATTAGCYGTGC CATATAATATGAGAGTTATCCATTTTGGTGCTGGCTCRGACAAAGG AGTTGCACCCGGCACAGCTGTTCTCAGACAGTGGTTGCCAATTG GCACACTACTTGTTGATTGAGATCTTAACGACTTCGTCTCTGACG CTGATTCCACTCTAATTGGAGACTGTGCAACCGTACATACAGCTAA CAAATGGGATCTCATTATTAGCGATATGTATGATCCTAAAACCAAAC ACGTGACAAAGGAAAAATGATTCAAAAGAAGGATTTTCACTTACCT GTGTGGATTTATTAACAAAAAATTAGCCCTGGGAGGCTCTGTGGC TGTAAGATAACTGAGCATTCTTGGAATGCGGATCTCTACAAAGCTC ATGGGACATTTCTCATGGTGGACAGCTTTTGTACAAATGTTAATG CATCTTCATCAGAAGCATTTTAAATTGGAGTTAACTATCTTGGTAA CCAAAAGAACAAATTGATGGTTACACCATGCATGCTAACTACATTT TCTGGAGGAATACAAACCCGATTCAATTGTCTTCTTATTCACTTTT TGACATGAGTAAGTTCCCTCTTAATTAAGGGGAACAGCTGTCAT GTCTTTAAAGGAGAACCAAATCAATGAAATGATTTATTCTCTACTTG AAAAAGGCAGACTTATCATTAGGGAAAACAACAGAGTTGTTGTCT CAAGTGATGTTCTTGTTAACTAAACGAACA	
ORF3a	ATGGACCTGTTTCATGAGGATCTTCACCATCGGCACCGTGACCCT GAAGCAGGGCGAGATCAAGGACGCCACCCCAAGCGACTTCGTG AGGGCCACCGCCACCATCCCCATCCAGGCCAGCCTGCCCTTCG GCTGGCTGATCGTGGGCGTGGCCCTGCTGGCCGTGTTCCAGAG CGCCAGCAAGATCATCACCCCTGAAGAAGAGGTGGCAGCTGGCC CTGAGCAAGGGCGTGCACTTCGTGTGCAACCTGCTGCTGCTGTT CGTGACCGGTGTACAGCCACCTGCTGCTGGTGGCCGCCGGCCTG GAGGCCCCCTTCTGTACCTGTACGCCCTGGTGTACTTCTGTGCA GAGCATCAACTTCGTGAGGATCATCATGAGGCTGTGGCTGTGCT GGAAGTGCAAGGAGCAAGAACCCCTGCTGTACGACGCCAACTAC TTCTGTGCTGGCACACCAACTGCTACGACTACTGCATCCCCTAC	152

	AACAGCGTGACCAGCAGCATCGTGATCACCAGCGGCGACGGCA CCACCAGCCCCATCAGCGAGCAGACTACCAGATCGGCGGCTAC ACCGAGAAGTGGGAGAGCGGCGTGAAGGACTGCGTGGTGCTGC ACAGCTACTTCACCAGCGACTACTACCAGCTGTACAGCACCCAG CTGAGCACCGACACCGGCGTGGAGCACGTGACCTTCTTCATCTA CAACAAGATCGTGGACGAGCCCCGAGGAGCACGTGCAGATCCAC ACCATCGACGGCAGCAGCGGCGTGGTGAACCCCGTGATGGAGC CCATCTACGACGAGCCCACCACCACCAGCGTGCCCCTG	
Envelope (E)	ATGTACAGCTTCGTGAGCGAGGAGACCGGCACCCTGATCGTGAA CAGCGTGCTGCTGTTCTTGGCCTTCGTGGTGTTCCTGCTGGTGA CCCTGGCCATCCTGACCGCCCTGAGGCTGTGCGCCTACTGCTG CAACATCGTGAACGTGAGCCTGGTGAAGCCCAGCTTCTACGTGT ACAGCAGGGTGAAGAACCTGAACAGCAGCAGGGTGCCCCGACCT GCTGGTG	153
Membrane (M)	ATGGCCGACAGCAACGGCACCATCACCGTGGAGGAGCTGAAGA AGCTGCTGGAGCAGTGAACCTGGTGATCGGCTTCCTGTTCTCTG ACCTGGATCTGCCTGCTGCAGTTCGCCTACGCCAACAGGAACAG GTTCTGTACATCATCAAGCTGATCTTCTGTGGCTGCTGTGGCC CGTGACCCCTGGCCTGCTTCGTGCTGGCCGCGCGTGACAGGATCA ACTGGATCACCGGCGGCATCGCCATCGCCATGGCCTGCCTGGTG GGCCTGATGTGGCTGAGCTACTTCATCGCCAGCTTCAGGCTGTT CGCCAGGACCAGGAGCATGTGGAGCTTCAACCCCGAGACCAAC ATCCTGCTGAACGTGCCCCTGCACGGCACCATCCTGACCAGGCC CCTGCTGGAGAGCGAGCTGGTGATCGGCGCCGTGATCCTGAGG GGCCACCTGAGGATCGCCGCCACCACTGGGCAGGTGCGACA TCAAGGACCTGCCCCAAGGAGATCACCGTGGCCACCAGCAGGAC CCTGAGCTACTACAAGCTGGGCGCCAGCCAGAGGGTGGCCGGC GACAGCGGCTTCGCCGCTTACAGCAGGTACAGGATCGGCAACTA CAAGCTGAACACCGACCACAGCAGCAGCAGCGACAACATCGCC CTGCTGGTGACG	154
ORF6	ATGTTCCACCTGGTGGACTTCCAGGTGACCATCGCCGAGATCCT GCTGATCATCATGAGGACCTTCAAGGTGAGCATCTGGAACCTGG ACTACATCATCAACCTGATCATCAAGAACCTGAGCAAGAGCCTGA CCGAGAACAAGTACAGCCAGCTGGACGAGGAGCAGCCCATGGA GATCGAC	155
ORF7a	ATGAAGATCATCCTGTTCTTGGCCCTGATCACCTGGCCACCTGC GAGCTGTACCACTACCAGGAGTGCGTGAGGGGCACCACCGTG	156
Nsp3	GCCCCACCAAGGTGACCTTCGGCGACGACACCGTGATCGAGG TGCAGGGCTACAAGAGCGTGAACATCACCTTCGAGCTGGACGAG AGGATCGACAAGGTGCTGAACGAGAAGTGCAGCGCCTACACCGT GGAGCTGGGCACCGAGGTGAACGAGTTCGCCTGCGTGGTGGCC GACGCCGTGATCAAGACCTGCAGCCCGTGAGCGAGCTGCTGA CCCCCTGGGCATCGACCTGGACGAGTGGAGCATGGCCACCTA CTACCTGTTTCGACGAGAGCGGCGAGTTCAAGCTGGCCAGCCAC ATGTAATGCACTTCTACCCCCCGACGAGGACGAGGAGGAGG GCGACTGCGAGGAGGAGGAGTTCGAGCCCAGCACCCAGTACGA GTACGGCACCGAGGACGACTACCAGGGCAAGCCCCCTGGAGTTC GGCGCCACCAGCGCCGCCCTGCAGCCCAGGAGGAGCAGGAG GAGGACTGGCTGGACGACGACAGCCAGCAGACCGTGGGCCAG CAGGACGGCAGCGAGGACAACCAGACCACCACCATCCAGACCA TCGTGGAGGTGACGCCCCAGCTGGAGATGGAGCTGACCCCCGT GGTGCAGACCATCGAGGTGAACAGCTTCAGCGGCTACCTGAAGC TGACCGACAACGTGTACATCAAGAACGCCGACATCGTGGAGGAG	157

	GCCAAGAAGGTGAAGCCCACCGTGGTGGTGAACGCCGCCAACG TGTACCTGAAGCACGGCGGGCGGTGGCCGGCGCCCTGAACAA GGCCACCAACAACGCCATGCAGGTGGAGAGCGAGCTACATC GCCACCAACGGCCCCCTGAAGGTGGGCGGCAGCTGCGTGTGA GCGGCCACAACCTGGCCAAGCACTGCCTGCACGTGGTGGGCCC CAACGTGAACAAGGGCGAGGACATCCAGCTGCTGAAGAGCGCC TACGAGAACTTCAACCAGCACGAGGTGCTGCTGGCCCCCTGCT GAGCGCCGGCATCTTCGGCGCCGACCCCATCCACAGCCTGAGG GTGTGCGTGGACACCGTGAGGACCAACGTGTACCTGGCCGTGT TCGACAAGAACCTGTACGACAAGCTGGTGAAGCAGCTTCTGGAG ATGAAGAGCGAGAAGCAGGTGGAGCAGAAGATCGCCGAGATCC CCAAGGAGGAGGTGAAGCCCTTCATCACCGAGAGCAAGCCAG CGTGGAGCAGAGGAAGCAGGACGACAAGAAGATCAAGGCCTGC GTGGAGGAGGTGACCACCAACCTGGAGGAGACCAAGTTCTCTGA CCGAGAACCTGCTGCTGTACATCGACATCAACGGCAACCTGCAC CCCGACAGCGCCACCCCTGGTGAAGCAGATCGACATCACCTTCCT GAAGAAGGACGCCCCCTACATCGTGGGCGACGTGGTGCAGGAG GGCGTGTGACCGCCGTGGTGATCCCCACCAAGAAGGCCGGCG GCACCAACCGAGATGCTGGCCAAGGCCCTGAGGAAGGTGCCAC CGACAACCTACATCACCACTACCCCGGCCAGGGCCTGAACGGCT ACACCGTGGAGGAGGCCAAGACCGTGTGAAGAAGTGCAAGAG CGCCTTCTACATCCTGCCAGCATCATCAGCAACGAGAAGCAGG AGATCCTGGGCACCGTGAGCTGGAACCTGAGGGAGATGCTGGC CCACGCCGAGGAGACCAGGAAGCTGATGCCCCGTGTGCGTGGAG ACCAAGGCCATCGTGAGCACCATCCAGAGGAAGTACAAGGGCAT CAAGATCCAGGAGGGCGTGGTGACTACGGCGCCAGGTTCTACT TCTACACCAGCAAGACCACCGTGGCCAGCCTGATCAACACCCTG AACGACCTGAACGAGACCTGCTGACCATGCCCCCTGGGCTACGT GACCCACGGCCTGAACCTGGAGGAGGCCGCCAGGTACATGAGG AGCCTGAAGGTGCCCCGCCACCGTGAGCGTGAGCAGCCCCGACG CCGTGACCGCCTACAACGGCTACCTGACCAGCAGCAGCAAGAC CCCCGAGGAGCACTTCATCGAGACCATCAGCCTGGCCGGCAGC TACAAGGACTGGAGCTACAGCGGCCAGAGCACCCAGCTGGGCA TCGAGTTCTGAAGAGGGGCGACAAGAGCGTGACTACACCAGC AACCCACCACTTCCACCTGGACGGCGAGGTGATCACCTTCGA CAACCTGAAGACCCTGCTGAGCCTGAGGGAGGTGAGGACCATC AAGGTGTTCAACACCGTGGACAACATCAACCTGCAC	
Nsp5	AGCGGCTTCAGGAAGATGGCCTTCCCCAGCGGCAAGGTGGAGG GCTGCATGGTGCAGGTGACCTGCGGCCACCAACCCCTGAACGG CCTGTGGCTGGACGACGTGGTGTACTGCCCCAGGCACGTGATCT GCACCAAGCGAGGACATGCTGAACCCCAACTACGAGGACCTGCT GATCAGGAAGAGCAACCACAACCTTCTGGTGCAGGCCGGCAAC GTGCAGCTGAGGGTGATCGGCCACAGCATGCAGAACTGCGTGC TGAAGCTGAAGGTGGACACCGCCAACCCCAAGACCCCAAGTAC AAGTTCGTGAGGATCCAGCCCGGCCAGACCTTCAGCGTGCTGG CCTGCTACAACGGCAGCCCCAGCGGCGTGTACCAAGTGCGCCAT GAGGCCCAACTTACCATCAAGGGCAGCTTCTGAACGGCAGCT GCGGCAGCGTGGGCTTCAACATCGACTACGACTGCGTGAGCTTC TGCTACATGCACCACATGGAGCTGCCCACCGGCGTGACGCCG GCACCGACCTGGAGGGCAACTTCTACGGCCCCCTTCGTGGACAG GCAGACCGCCCAGGCCGCCGGCACCGACACCAACCATCACCGTG AACGTGCTGGCCTGGCTGTACGCCGCCGTGATCAACGGCGACA GGTGGTTCTGAACAGGTTTACCACCAACCTGAACGACTTCAAC CTGGTGGCCATGAAGTACAACCTACGAGCCCCCTGACCCAGGACCA CGTGGACATCCTGGGCCCCCTGAGCGCCAGACCGGCATCGCC GTGCTGGACATGTGCGCCAGCCTGAAGGAGCTGCTGCACAACG GCATGAACGGCAGGACCATCCTGGGCAGCGCCCTGCTGGAGGA	158

	CGAGTTCACCCCCTTCGACGTGGTGAGGCAGTGCAGCGGCGTG ACCTTCCAG	
Nsp12	AGCGCCGACGCCAGAGCTTCTGAACAGGGTGTGCGGCGTGA GCGCCGCCAGGCTGACCCCCTGCGGCACCGGCACCAGCACCG ACGTGGTGTACAGGGCCTTCGACATCTACAACGACAAGGTGGCC GGCTTCGCCAAGTTCTGAAGACCAACTGCTGCAGGTTCCAGGA GAAGGACGAGGACGACAACCTGATCGACAGCTACTTCGTGGTGA AGAGGCACACCTTCAGCAACTACCAGCACGAGGAGACCATCTAC AACCTGCTGAAGGACTGCCCCGCGTGGCCAAGCACGACTTCTT CAAGTTCAGGATCGACGGCGACATGGTGCCCCACATCAGCAGGC AGAGGCTGACCAAGTACACCATGGCCGACCTGGTGTACGCCCTG AGGCACTTCGACGAGGGCAACTGCGACACCCTGAAGGAGATCC TGGTGACCTACAACCTGCTGCGACGACGACTACTTCAACAAGAAG GACTGGTACGACTTCGTGGAGAACCCCGACATCCTGAGGGTGT CGCCAACCTGGGCGAGAGGGTGAGGCAGGCCCTGCTGAAGACC GTGCAGTTCTGCGACGCCATGAGGAACGCCGGCATCGTGGGCG TGCTGACCCTGGACAACCAGGACCTGAACGGCAACTGGTACGAC TTCGGCGACTTCATCCAGACCACCCCGGCAGCGGCGTGCCCG TGGTGGACAGCTACTACAGCCTGCTGATGCCCATCCTGACCCTG ACCAGGGCCCTGACCGCCGAGAGCCACGTGGACACCGACCTGA CCAAGCCCTACATCAAGTGGGACCTGCTGAAGTACGACTTCACC GAGGAGAGGCTGAAGCTGTTGACAGGTACTTCAAGTACTGGGA CCAGACCTACCACCCCAACTGCGTGAAGTGCCTGGACGACAGGT GCATCCTGCACTGCGCCAACCTCAACGTGCTGTTACGACCCGTG TTCCCCCCCCACAGCTTCGGCCCCCTGGTGAGGAAGATCTTCGT GGACGGCGTGCCCTTCGTGGTGAGCACCGGCTACCACTTCAGG GAGCTGGGCGTGGTGACAAACCAGGACGTGAACCTGCACAGCA GCAGGCTGAGCTTCAAGGAGCTGCTGGTGTACGCCGCCGACCC CGCCATGCACGCCGCCAGCGGCAACCTGCTGCTGGACAAGAGG ACCACCTGCTTCAGCGTGGCCGCCCTGACCAACAACGTGGCCTT CCAGACCGTGAAGCCCCGCAACTTCAACAAGGACTTCTACGACT TCGCCGTGAGCAAGGGCTTCTTCAAGGAGGGCAGCAGCGTGA GCTGAAGCACTTCTTCTCGCCCAGGACGGCAACGCCGCCATCA GCGACTACGACTACTACAGGTACAACCTGCCACCATGTGCGAC ATCAGGCAGCTGCTGTTCTGGTGGAGGTGGTGGACAAGTACTT CGACTGCTACGACGGCGGCTGCATCAACGCCAACAGGTGATCG TGAACAACCTGGACAAGAGCGCCGGCTTCCCTTCAACAAGTGG GGCAAGGCCAGGCTGTACTACGACAGCATGAGCTACGAGGACCA GGACGCCCTGTTGCGCTACACCAAGAGGAACGTGATCCCCACCA TCACCCAGATGAACCTGAAGTACGCCATCAGCGCCAAGAACAGG GCCAGGACCGTGGCCGGCGTGAGCATCTGCAGCACCATGACCA ACAGGCAGTTCACACAGAAGCTGCTGAAGAGCATCGCCGCCAC CAGGGGCGCCACCGTGGTGTATCGGCACCAGCAAGTTCTACGGC GGCTGGCACAAACATGCTGAAGACCGTGTACAGCGACGTGGAGA ACCCCCACCTGATGGGCTGGGACTACCCCAAGTGGCAGAGGGC CATGCCCAACATGCTGAGGATCATGGCCAGCCTGGTGTGGCCA GGAAGCACACCACCTGCTGCAGCCTGAGCCACAGGTTCTACAG GCTGGCCAACGAGTGCGCCAGGTGCTGAGCGAGATGGTGTATG TGCGGCGGCAGCCTGTACGTGAAGCCCCGGCGGCACCAGCAGC GGCGACGCCACCACCGCCTACGCCAACAGCGTGTTCACATCTG CCAGGCCGTGACCGCCAACGTGAACGCCCTGCTGAGCACCGGAC GGCAACAAGATCGCCGACAAGTACGTGAGGAACCTGCAGCACA GGCTGTACGAGTGCCTGTACAGGAACAGGGACGTGGACACCGA CTTCGTGAACGAGTTCTACGCCCTACCTGAGGAAGCACTTCAGCAT GATGATCCTGAGCGACGACGCCGTGGTGTGCTTCAACAGCACCT ACGCCAGCCAGGGCCTGGTGGCCAGCATCAAGAACTTCAAGAG CGTGCTGTACTACCAGAACAACGTGTTTATGAGCGAGGCCAAGT	159

	GCTGGACCGAGACCGACCTGACCAAGGGCCCCACGAGTTCTG CAGCCAGCACACCATGCTGGTGAAGCAGGGCGACGACTACGTG TACCTGCCCTACCCGACCCAGCAGGATCCTGGGCGCCGGCT GCTTCGTGGACGACATCGTGAAGACCGACGGCACCCTGATGATC GAGAGGTTCTGTGAGCCTGGCCATCGACGCCTACCCCTGACCAA GCACCCAACCAGGAGTACGCCGACGTGTTCCACCTGTACCTGC AGTACATCAGGAAGCTGCACGACGAGCTGACCGGCCACATGCTG GACATGTACAGCGTGATGCTGACCAACGACAACACCAGCAGGTA CTGGGAGCCCGAGTTCTACGAGGCCATGTACACCCCCACACCG TGCTGCAG	
--	--	--

Molecular Adjuvants and T Cell Enhancements

[00279] In certain embodiments, the vaccine composition comprises a molecular adjuvant and/or one or more T Cell enhancement compositions. The adjuvant and/or enhancement compositions may help improve the immunogenicity and/or long-term memory of the vaccine composition. Non-limiting examples of molecular adjuvants include CpG, such as a CpG polymer, and flagellin.

[00280] In some embodiments, the vaccine composition comprises a T cell attracting chemokine. The T cell attracting chemokine helps pull the T cells from the circulation to the appropriate tissues, e.g., the lungs, heart, kidney, and brain. Non-limiting examples of T cell attracting chemokines include CCL5, CXCL9, CXCL10, CXCL11, CCL25, CCL28, CXCL14, CXCL17, or a combination thereof.

[00281] In some embodiments, the vaccine composition comprises a composition that promotes T cell proliferation. Non-limiting examples of compositions that promote T cell proliferation include IL-7, IL-15, IL-2, or a combination thereof.

[00282] In some embodiments, the vaccine composition comprises a composition that promotes T cell homing in the lungs. Non-limiting examples of compositions that promote T cell homing include CCL25, CCL28, CXCL14, CXCL17 or a combination thereof.

[00283] In certain embodiments, the molecular adjuvant and/or the T cell attracting chemokine and/or the composition that promotes T cell proliferation are delivered with a separate antigen delivery system from the large sequences.

[00284] Table 11 shows non-limiting examples of T-cell enhancements that may be used to create a vaccine composition described herein.

Table 11

T-cell enhancement	Sequence	SEQ ID NO:
CXCL11	ATGAACAGGAAGGTGACCGCCATCGCCCTGGCCGCCATCATCTGGG CCACCGCCGCCAGGGCTTCCTGATGTTCAAGCAGGGCAGGTGCC TGTGCATCGGCCCGGCATGAAGGCCGTGAAGATGGCCGAGATCGA GAAGGCCAGCGTGATCTACCCAGCAACGGCTGCGACAAGGTGGA	160

	GGTGATCGTGACCATGAAGGCCCAAGAGGCAGAGGTGCCTGGA CCCCAGGAGCAAGCAGGCCAGGCTGATCATGCAGGCCATCGAGAA GAAGAACTTCCTGAGGAGGCAGAACATGTGA	
CCL5	ATGAAGGTCTCCGCGGCAGCCCTCGCTGTCATCCTCATTGCTACTGC CCTCTGCGCTCCTGCATCTGCCTCCCCATATTCTCGGACACCACAC CCTGCTGCTTTGCCTACATTGCCCGCCCACTGCCCCGTGCCACAT CAAGGAGTATTTCTACACCAAGTGGCAAGTGTCCAACCCAGCAGTC GTCCACAGGTCAAGGATGCCAAAGAGAGAGGGACAGCAAGTCTGG CAGGATTTCTGTATGACTCCCGGTGAACAAGGGCAAGCTTTGTCA CCCGAAAGAACCGCCAAGTGTGTGCCAACCAGAGAAGAAATGGGT TCGGGAGTACATCAACTCTTTGGAGATGAGCTAGGATGGAGAGTCCT TGAACCTGAACTTACACAAATTTGCCTGTTTCTGCTTGCTCTTGTCT AGCTTGGGAGGCTTCCCTCACTATCCTACCCACCCGCTCCTTGA	161
CXCL9	ATGAAGAAAAGTGGTGTTCTTTTCTTGGGCATCATCTTGCTGGTT CTGATTGGAGTGCAAGGAACCCCAAGTAGTGAGAAAGGGTCGCTGTT CCTGCATCAGCACCAACCAAGGACTATCCACCTACAATCCTTGAA GACCTTAAACAATTTGCCCAAGCCCTTCTGCGAGAAAATTGAAAT CATTGCTACACTGAAGAATGGAGTTCAAACATGTCTAAACCCAGATTC AGCAGATGTGAAGGAACTGATTAAAAAGTGGGAGAAACAGGTCAGC CAAAAGAAAAAGCAAAAGAATGGGAAAAAACATCAAAAAAGAAAGT TCTGAAAGTTGAAAATCTCAACGTTCTCGTCAAAAGAAGACTACATA A	162
CXCL10	ATGAATCAAACCTGCCATTCTGATTTGCTGCCTTATCTTTCTGACTCTAA GTGGCATTCAAGGAGTACCTCTCTCTAGAACTGTACGCTGTACCTGC ATCAGCATTAGTAATCAACCTGTTAATCCAAGGTCTTTAGAAAACTTG AAATTATTCCTGCAAGCCAATTTTGTCCACGTGTTGAGATCATTGCTA CAATGAAAAAGAAGGGTGAGAAGAGATGTCTGAATCCAGAATCGAAG GCCATCAAGAATTTACTGAAAGCAGTTAGCAAGGAAAGGTCTAAAG ATCTCCTTAA	163
CXCL14	ATGAGGCTCCTGGCGGCCGCGCTGCTCCTGCTGCTGCTGGCGCTG TACACCGCGCGTGTGGACGGGTCCAAATGCAAGTGCTCCCGGAAG GGACCCAAGATCCGCTACAGCGACGTGAAGAAGCTGGAATGAAGC CAAAGTACCCGCACTGCGAGGAGAAGATGTTATCATCACCACCAAG AGCGTGTCCAGGTACCGAGGTCAGGAGCACTGCCTGCACCCCAAG CTGCAGAGCACCAAGCGCTTCATCAAGTGGTACAACGCCTGGAACG AGAAGCGCAGGGTCTACGAAGAATAG	164
CXCL17	ATGAAAGTTCTAATCTCTTCCCTCCTCCTGTTGCTGCCACTAATGCTG ATGTCCATGGTCTCTAGCAGCCTGAATCCAGGGGTGCGCCAGAGGCC ACAGGGACCGAGGCCAGGCTTCTAGGAGATGGCTCCAGGAAGGCG GCCAAGAATGTGAGTGCAAAGATTGGTTCTGAGAGCCCCGAGAAG AAAATTCATGACAGTGTCTGGGCTGCCAAGAAGCAGTGCCCCTGT GATCATTTCAAGGGCAATGTGAAGAAAACAAGACACCAAAGGCACCA CAGAAAGCCAAACAAGCATTCCAGAGCCTGCCAGCAATTTCTCAAAC AATGTCAGCTAAGAAGCTTTGCTCTGCCTTTGTAG	165
CCL25	ATGAACCTGTGGCTCCTGGCCTGCCTGGTGGCCGGCTTCCTGGGA GCCTGGGCCCCCGCTGTCCACACCCAAGGTGTCTTTGAGGACTGCT GCCTGGCCTACCACTACCCATTGGGTGGGCTGTGCTCCGGCGCG CCTGGACTTACCGGATCCAGGAGGTGAGCGGGAGCTGCAATCTGCC TGCTGCGATATTCTACCTCCCCAAGAGACACAGGAAGGTGTGTGGG AACCCAAAAGCAGGGAGGTGCAGAGAGCCATGAAGCTCCTGGATG CTCGAAATAAGGTTTTTCAAAGCTCCACCACAACACGCAGACCTTC CAAGCAGGCCCTCATGCTGTAAGAAGTTGAGTTCTGGAACTCCAA	166

	GTTATCATCGTCCAAGTTTAGCAATCCCATCAGCAGCAGTAAGAGGA ATGTCTCCCTCCTGATATCAGCTAATTCAGGACTGTGA	
CCL28	ATGCAGCAGAGAGGACTCGCCATCGTGGCCTTGGCTGTCTGTGCGG CCCTACATGCCTCAGAAGCCATACTTCCCATTGCCTCCAGCTGTTGC ACGGAGGTTTTACATCATATTTCCAGAAGGCTCCTGGAAAGAGTGAA TATGTGTCGCATCCAGAGAGCTGATGGGGATTGTGACTTGGCTGCTG TCATCCTTCATGTCAAGCGCAGAAGAATCTGTGTGAGCCCGCACAAAC CATACTGTTAAGCAGTGGATGAAAGTGCAAGCTGCCAAGAAAAATGG TAAAGGAAATGTTTGCCACAGGAAGAAACACCATGGCAAGAGGAAC AGTAACAGGGCACATCAGGGGAAACACGAAACATACGGCCATAAAAC TCCTTATTAG	167
IL-7	ATGTTCCACGTGAGCTTCAGGTACATCTTCGGCATCCCCCCTGAT CCTGGTGCTGCTGCCCCGTGACCAGCAGCGAGTGCCACATCAAGGA CAAGGAGGGCAAGGCCTACGAGAGCGTGCTGATGATCAGCATCGAC GAGCTGGACAAGATGACCGGCACCGACAGCAACTGCCCAACAAC GAGCCCAACTTCTTCAGGAAGCAGTGTGCGACGACACCAAGGAG GCCGCCTTCCTGAACAGGGCCGCCAGGAAGCTGAAGCAGTTCCTG AAGATGAACATCAGCGAGGAGTTCAACGTGCACCTGCTGACCGTGA GCCAGGGCACCCAGACCCTGGTGAAGTGCACCAGCAAGGAGGAGA AGAAGCTGAAGGAGCAGAAGAAGAAGACGCGCTGCTTCCTGAAGA GGCTGCTGAGGGAGATCAAGACCTGCTGGAACAAGATCCTGAAGGG CAGCATCTGA	168
IL-15	ATGAGAATTTGAAACCACATTTGAGAAGTATTTCCATCCAGTGCTAC TTGTGTTTACTTCTAAACAGTCATTTTCTAACTGAAGCTGGCATTGATG TCTTCATTTTGGGCTGTTTCAGTGACAGGGCTTCCTAAACAGAAGCC AACTGGGTGAATGTAATAAGTGATTTGAAAAAATTGAAGATCTTATTC AATCTATGCATATTGATGCTACTTTATATACGGAAAGTGATGTTCAACC CAGTTGCAAAGTAACAGCAATGAAGTGCTTTCTCTTGGAGTTACAAG TTATTTCACTTGAGTCCGGAGATGCAAGTATTCATGATACAGTAGAAA ATCTGATCATCCTAGCAAACAACAGTTTGTCTTCTAATGGGAATGTAA CAGAATCTGGATGCAAAGAATGTGAGGAACTGGAGGAAAAAATATT AAAGAATTTTTGCAGAGTTTTGTACATATTGTCCAAATGTTTCATCAACA CTTCTTGA	169
IL-2	ATGTACAGGATGCAACTCCTGTCTTGCACTTAAGTCTTGCACTT GTCACAAACAGTGCACCTACTTCAAGTTCTACAAAGAAAACACAGCT ACAACCTGGAGCATTTACTGCTGGATTACAGATGATTTTGAATGGAAT TAATAATTACAAGAATCCCAAACCTCACCAGGATGCTCACATTTAAGTTT TACATGCCCCAAGAAGGCCACAGAAGTGAACATCTTCAGTGTCTAGA AGAAGAAGTCAAACCTCTGGAGGAAGTGCTAAATTTAGCTCAAAGCA AAAACCTTCACTTAAGACCCAGGGACTTAATCAGCAATATCAACGTAA TAGTTCTGGAAGTAAAGGGATCTGAAACAACATTCATGTGTGAATATG CTGATGAGACAGCAACCATTGTAGAATTTCTGAACAGATGGATTACCT TTTGTCAAAGCATCATCTCAACACTGACTTGA	170

[00285] In preferred embodiments, the T-cell enhancement compositions described herein (e.g. CXCL9, CXCL10, IL-7, IL-2) may be integrated into a separate delivery system from the vaccine compositions. In some embodiments, the T-cell enhancement compositions described herein (e.g. CXCL9, CXCL10, IL-7, IL-2) may be integrated into a the same delivery system as the vaccine compositions.

[00286] In certain embodiments, the vaccine composition comprises a tag. For example, in some embodiments, the vaccine composition comprises a His tag. The present invention is not limited to a His

tag and includes other tags such as those known to one of ordinary skill in the art, such as a fluorescent tag (e.g., GFP, YFP, etc.), etc.

Antigen Delivery System

[00287] The present invention also features vaccine compositions in the form of an antigen delivery system. Any appropriate antigen delivery system may be considered for delivery of the antigens described herein. The present invention is not limited to the antigen delivery systems described herein.

[00288] In certain embodiments, the antigen delivery system is for targeted delivery of the vaccine composition, e.g., for targeting to the tissues of the body where the virus replicates.

[00289] In certain embodiments, the antigen delivery system comprises adenoviruses such as but not limited to Ad5, Ad26, Ad35, etc., as well as carriers such as lipid nanoparticles, polymers, peptides, etc. In other embodiments, the antigen delivery system comprises a vesicular stomatitis virus (VSV) vector.

[00290] The present invention is not limited to adenovirus vector-based antigen delivery systems. In certain embodiments, the antigen delivery system comprises an adeno-associated virus vector-based antigen delivery system, such as but not limited to the adeno-associated virus vector type 9 (AAV9 serotype), AAV type 8 (AAV8 serotype), etc. In certain embodiments, the adeno-associated virus vectors used are tropic, e.g., tropic to lungs, brain, heart and kidney, e.g., the tissues of the body that express ACE2 receptors (FIG. 3A)). For example, AAV9 is known to be neurotropic, which would help the vaccine composition to be expressed in the brain.

[00291] In the antigen delivery system, the one or more large sequences are operatively linked to a promoter. In certain embodiments, the one or more large sequences are operatively linked to a generic promoter. For example, in certain embodiments, the one or more large sequences are operatively linked to a CMV promoter. In certain embodiments, the one or more large sequences are operatively linked to a CAG, EF1A, EFS, CBh, SFFV, MSCV, mPGK, hPGK, SV40, UBC, or other appropriate promoter.

[00292] In some embodiments, the one or more large sequences are operatively linked to a tissue-specific promoter (e.g., a lung-specific promoter). For example, the antigen may be operatively linked to a SpB promoter or a CD144 promoter.

[00293] As discussed, in certain embodiments, the vaccine composition comprises a molecular adjuvant. In certain embodiments, the molecular adjuvant is operatively linked to a generic promoter, e.g., as described above. In certain embodiments, the molecular adjuvant is operatively linked to a tissue-specific promoter, e.g., a lung-specific promoter, e.g., SpB or CD144.

[00294] As discussed, in certain embodiments, the vaccine composition comprises a T cell attracting chemokine. In certain embodiments, the T cell attracting chemokine is operatively linked to a generic promoter, e.g., as described above. In certain embodiments, the T cell attracting chemokine is operatively linked to a tissue-specific promoter, e.g., a lung-specific promoter, e.g., SpB or CD144.

[00295] As discussed, in certain embodiments, the vaccine composition comprises a composition for promoting T cell proliferation. In certain embodiments, the composition for promoting T cell proliferation is operatively linked to a generic promoter, e.g., as described above. In certain embodiments, the composition for promoting T cell proliferation is operatively linked to a tissue-specific promoter, e.g., a lung-specific promoter, e.g., SpB or CD144.

[00296] Table 12 shows non-limiting examples of promoters that may be used to create a vaccine composition described herein.

Table 12

Promoter	Sequence	SEQ ID NO:
CAG	CTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT AGTTCATAGCCCATATATGGAGTTCGCGTTACATAACTTACGGTAAATGG CCCGCTGGCTGACCGCCCAACGACCCCGCCATTGACGTCAATAATG ACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATG GGTGGAGTATTACGGTAACTGCCACTTGGCAGTACATCAAGTGATC ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC TGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTA CATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGT TCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCAATTTGTAT TTATTTATTTTTAATTATTTTGTGCAGCGATGGGGCGGGGGGGGGGGG GGGGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGG GGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCC GAAAGTTTCCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAA GCGAAGCGCGCGGCGGGCGGGAGTTCGCTGCGCGCTGCCCTCGCCCC GTGCCCCGCTCCGCGCCGCTCGCGCCGCGCCGCGCCGCGCTCTGACT GACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCC GGGCTGTAATTAGCGCTTGGTTTAAATGACGGCTTGTTCCTTTCTGTGGC TGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGG AGCGGCTCGGGGGGTGCGTGCGTGTGTGTGCGTGCGGGAGCGCCGC GTGCGGCTCCGCGCTGCCGCGCGGCTGTGAGCGCTGCGGGCGCGG GCGGGGCTTTGTGCGCTCCGCAGTGTGCGCGAGGGGAGCGCGGCCGG GGGCGGTGCCCCGCGGTGCGGGGGGGGCTGCGAGGGGAACAAAGGC TGCGTGCGGGGTGTGTGCGTGCGGGGGGTGAGCAGGGGGTGTGGGCG CGTCCGTCCGGCTGCAACCCCCCTGCACCCCCCTCCCCAGTTGCTG AGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTACGGGGCGTGGCGCG GGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGG GCGGGGCGGGGCGCCCTCGGGCCGGGAGGGCTCGGGGGAGGGGC GCGGCGGCCCCCGGAGCGCCGCGGCTGTGAGGCGCGGCGAGCCG CAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCT TTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAGGCGCCGCCGCCACC CCCTCTAGCGGGCGGGGCGAAGCGGTGCGGCGCCGGCAGGAAGG AAATGGGCGGGGAGGGCCTTCGTGCGTCCGCGCGCCGCGCTCCCTT CTCCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTTCG GGGGGACGGGGCAGGGCGGGGTTCCGGCTTCTGGCGTGTGACCGGC GGCTCTAGAGCCTCTGCTAACCATGTTTCATGCCTTCTTTTCTTCTACAG CTCCTGGGCAACGTGCTGTTATTGTGCTGTCTCATCATTTTGGCAAAGA ATTG	171
CMV	TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATG GAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGC	172

	CCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTA ACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTACGGTA AACTGCCCACCTTGGCAGTACATCAAGTGATCATATGCCAAGTACGCCCC CTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCAGTAC ATGACCTTATGGGACTTTCCCTACTTGGCAGTACATCTACGTATTAGTCATC GCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATA GCGGTTTGA CTACG GGGATTTC AAGTCTCCACCCCATTGACGTCAAT GGGAGTTTGT TTTGGCACC AAAATCAACGGGACTTTCCAAAATGTCGTAA CAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGGTACGGTGGGA GGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATC	
SP-B	GTATAGGGCTGTCTGGGAGCCACTCCAGGGCCACAGAAATCTTGTCTCT GACTCAGGGTATTTTGT TTTCTGTTTTGTGTAAATGCTCTTCTGACTAATG CAAACCATGTGTCCATAGAACCAGAAGATTTTTCCAGGGGAAAAGGTAAG GAGGTGGTGAGAGTGTCTGGGTCTGCCCTTCCAGGGCTTGGCCTGGG TTAAGAGCCAGGCAGGAAGCTCTCAAGAGCATTGCTCAAGAGTAGAGGG GGCCTGGGAGGCCCAGGGAGGGGATGGGAGGGGAACACCCAGGCTG CCCCAACCCAGATGCCCTCCACCCTCCTCAACCTCCCTCCACGGCCT GGAGAGGTGGGACCAGGTATGGAGGCTTGAGAGCCCCCTGGTTGGAGGA AGCCACAAGTCCAGGAACATGGGAGTCTGGGCAGGGGGCAAAGGAGG CAGGAACAGGGCCATCAGCCAGGACAGGTGGTAAGGCAGGCAGGAGTGT TCCTGCTGGGAAAAGGTGGGATCAAGCACCTGGAGGGCTCTTCAGAGC AAAGACAAACACTGAGGTGCTGCTGCCACTCCTACAGAGCCCCACGCC CGCCCAGCTATAAGGGGCCATGCACCAAGCAGGGTACCCAGGCTGCAG AGGTGCC	173
CD144	CATCCATGCCCATGGCCTCAGATGCCAGCCATAAGCTGTTGGGTTCCAA ACCTCGACTCCAGGCTGGACTCACCCCTGTCTCCCCCACCAGCCTGAC ACCTCCACCTGGGTATCTAACGAGCATCTCAAACCTCAACCTGCCTGAGA CAGAGGAATCACTATCCCCTCCTCCTCCAAAAATATCCTTCCATCACACTC CCCATCTTGTGCTCTGATTACTAAACGGCCCTGGGCCCCTCTCTTTCTCA GGGTCTCTGCTTGCCAGCTATATAATAAAACAAGTTTGGGACTTCCCAA CCATTACCCCATGGAAAAACAGAAGCAACTCTTCAAAGGACAGATTCCTCA GGATCTGCCCTGGGAGATTCCAAATCAGTTGATCTGGGGTGAGCCCAGT CCTCTGTAGTTTTTAGAAGCTCCTCCTATGTCTCTCCTGGTCAGCAGAAT CTTGGCCCCCTCCCCTCCCCCAGCCTCTTGGTTCTTCTGGGCTCTGATC CAGCCTCAGCGTCACTGTCTTCCAGCCCCCTCTTTGATTCTCGTTTATGT CAAAAGCCTTGTGAGGATGAGGCTGTGATTATCCCCATTTTACAGATGAG GAACTGTGGCTCCAGGATGACACAACCTGGCCAGAGGTACATCAGAAAG CAGAGCTGGGTCACTTGACTCCACCCAATATCCCTAAATGCAAACATCCC CTACAGACCCGAGGCTGGCACCTTAGAGCTGGAGTCCATGCCCGCTCTG ACCAGGAGAAGCCAACCTGGTCCTCCAGAGCCAAGAGCTTCTGTCCCTT TCCCATCTCCTGAAGCCTCCCTGTCACCTTTAAAGTCCATTCCACAAAAG ACATCATGGGATCACACAGAAAATCAAGCTCTGGGGCTAGGCTGACCC CAGCTAGATTTTTGGCTCTTTTATACCCAGCTGGGTGGACAAGCACCTT AAACCCGCTGAGCCTCAGCTTCCCGGGCTATAAAATGGGGGTGATGACA CCTGCCGTAGCATTCCAAGGAGGGTTAAATGTGATGCTGCAGCCAAGG GTCCCCACAGCCAGGCTCTTTCAGGTGCTGGGTTTCAGAGTCCCAGAG CTGAGGCCGGGAGTAGGGTTCAAGTGGGGTGCCCCAGGCAGGGTCC AGTGCCAGCCCTCTGTGGAGACAGCCATCCGGGGCCGAGGCAGCCGC CCACCGCAGGGCCTGCCTATCTGCAGCCAGCCCAGCCCTCACAAAGGA ACAATAACAGGAAACCATCCCAGGGGGAAGTGGGCCAGGGCCAGCTGG AAAACCTGAAGGGGAGGCAGCCAGGCCTCCCTCGCCAGCGGGGTGTG GCTCCCCTCCAAAGACGGTCGGCTGACAGGCTCCACAGAGCTCCACTC ACGCTCAGCCCTGGACGGACAGGCAGTCCAACGGAAACAGAAACATCCC TCAGCCCACAGGCAGGTGAGTGGGGGCTCCACACTCCCCTCCACCC CAAACCCGCCACCCTGCGCCCAAGATGGGAGGGTCCCTCAGCTTCCCA TCTGTAGAATGGGCATCGTCCCACTCCCATGACAGAGAGGCTCC	174

wild type native leader sequence	ATGTTTCGTGTTCTGGTGCTGCTGCCCCTGGTGAGCAGC	175
---	---	-----

[00297] In certain embodiments, the T cell attracting chemokine and the composition that promotes T cell proliferation are driven by the same promoter (e.g., the T cell attracting chemokine and the composition that promotes T cell proliferation are synthesized as a peptide). In certain embodiments, the T cell attracting chemokine and the composition that promotes T cell proliferation are driven by different promoters. In certain embodiments, the antigen, the T cell attracting chemokine, and the composition that promotes T cell proliferation are driven by the same promoter. In certain embodiments, the antigen, the T cell attracting chemokine, and the composition that promotes T cell proliferation are driven by the different promoters. In certain embodiments, the T cell attracting chemokine and the composition that promotes T cell proliferation are driven by the same promoter, and the one or more large sequences are driven by a different promoter.

[00298] In some embodiments, the antigen delivery system comprises one or more linkers between the T cell attracting chemokine and the composition that promotes T cell proliferation. In certain embodiments, linkers are used between one or more of the epitopes. The linkers may allow for cleavage of the separate molecules (e.g., chemokine). For example, in some embodiments, a linker is positioned between IL-7 (or IL-2) and CCL5, CXCL9, CXCL10, CXCL11, CCL25, CCL28, CXCL14, CXCL17, etc. In some embodiments, a linker is positioned between IL-15 and CCL5, CXCL9, CXCL10, CXCL11, CCL25, CCL28, CXCL14, CXCL17, etc. In some embodiments, a linker is positioned between the antigen or large sequence and another composition, e.g., IL-15, IL-7, CCL5, CXCL9, CXCL10, CXCL11, CCL25, CCL28, CXCL14, CXCL17, etc. A non-limiting example of a linker is T2A, E2A, P2A (see Table 13), or the like. The composition may feature a different linker between each open reading frame.

Table 13:

	SEQUENCE	SEQ ID NO:
T2A Linker	GGAAGCGGAGAGGGCAGGGGAAGTCTTCTAACATGCGGGGACGTGG AGGAAAATCCCGGCCCC	176
E2A Linker	GGAAGCGGACAGTGTACTAATTATGCTCTCTTGAAATTGGCTGGAGAT GTTGAGAGCAACCCAGGTCCC	177
P2A Linker	GGAAGCGGAGCCACGAACCTTCTCTGTAAAGCAAGCAGGAGATGT TGAAGAAAACCCCGGGCCT	178
	GCCGCCTAC	179
	GGCCCCGGCCCCGGC	180

6-His Tag	CATCACCATCACCATCAC	181
-----------	--------------------	-----

[00299] The present invention includes mRNA sequences encoding any of the vaccine compositions or portions thereof herein, e.g., a molecular adjuvant, a T cell enhancement, etc. The present invention also includes modified mRNA sequences encoding any of the vaccine compositions or portions thereof herein. The present invention also includes DNA sequence encoding any of the vaccine compositions or portions thereof herein.

[00300] In certain embodiments, nucleic acids of a vaccine composition herein are chemically modified. In some embodiments, the nucleic acids of a vaccine composition therein are unmodified. In some embodiments, all or a portion of the uracil in the open reading frame has a chemical modification. In some embodiments, a chemical modification is in the 5-position of the uracil. In some embodiments, a chemical modification is a N1-methyl pseudouridine. In some embodiments, all or a portion of the uracil in the open reading frame has a N1-methyl pseudouridine in the 5-position of the uracil.

[00301] In certain embodiments, an open reading frame of a vaccine composition herein encodes one antigen or epitopes. In some embodiments, an open reading frame of a vaccine composition herein encodes two or more antigens or epitopes. In some embodiments, an open reading frame of a vaccine composition herein encodes five or more antigens or epitopes. In some embodiments, an open reading frame of a vaccine composition herein encodes ten or more antigens or epitopes. In some embodiments, an open reading frame of a vaccine composition herein encodes 50 or more antigens or epitopes.

Methods

[00302] In some embodiments, the method comprises determining one or more conserved large sequences that are derived from coronavirus sequences (e.g., SARS-CoV-2, variants, common cold coronaviruses, previously known coronavirus strains, animal coronaviruses, etc.). The method may comprise selecting at least one large conserved sequence and synthesizing an antigen (or antigens) comprising the selected large conserved sequence(s). The method may comprise synthesizing a nucleotide composition (e.g., DNA, modified DNA, mRNA, modified mRNA, antigen delivery system, etc.) encoding the antigen comprising the selected large conserved sequence(s). In some embodiments, the method further comprises creating a vaccine composition comprising the antigen, nucleotide compositions, and/or antigen delivery system and a pharmaceutical carrier. In some embodiments, the large sequences comprise one or more conserved epitopes described herein, e.g., one or more conserved B-cell target epitopes and/or one or more conserved CD4+ T cell target epitopes and/or one or more conserved CD8+ T cell target epitopes.

[00303] In some embodiments, each of the large sequences are conserved among two or a combination of: at least two SARS-CoV-2 human strains in current circulation, at least one *coronavirus* that has caused a previous human outbreak, at least one *coronavirus* isolated from bats, at least one *coronavirus* isolated from pangolin, at least one *coronavirus* isolated from civet cats, at least one *coronavirus* strain isolated

from mink, and at least one *coronavirus* strain isolated from camels or any other animal that is receptive to *coronavirus*.

[00304] As previously discussed, the compositions described herein, e.g., the antigens, the vaccine compositions, the antigen delivery systems, the chemokines, the adjuvants, etc. may be used to prevent a *coronavirus* disease in a subject. In some embodiments, the compositions described herein, e.g., the antigens, the vaccine compositions, the antigen delivery systems, the chemokines, the adjuvants, etc. may be used to prevent a *coronavirus* infection prophylactically in a subject. In some embodiments, the compositions described herein, e.g., the antigens, the vaccine compositions, the antigen delivery systems, the chemokines, the adjuvants, etc. may elicit an immune response in a subject. In some embodiments, the compositions described herein, e.g., the antigens, the vaccine compositions, the antigen delivery systems, the chemokines, the adjuvants, etc. may prolong an immune response induced by the multi-epitope pan-*coronavirus* vaccine composition and increases T-cell migration to the lungs.

[00305] Methods for preventing a coronavirus disease in a subject may comprise administering to the subject a therapeutically effective amount of a pan-coronavirus vaccine composition according to the present invention. In some embodiments, the composition elicits an immune response in the subject. In some embodiments, the composition induces memory B and T cells. In some embodiments, the composition induces resident memory T cells (Trm). In some embodiments, the composition prevents virus replication, e.g., in the areas where the virus normally replicates such as lungs, brain, heart, and kidney. In some embodiments, the composition prevents a cytokine storm, e.g., in the areas where the virus normally replicates such as lungs, brain, heart, and kidney. In some embodiments, the composition prevents inflammation or an inflammatory response, e.g., in the areas where the virus normally replicates such as lungs, brain, heart, and kidney. In some embodiments, the composition improves homing and retention of T cells, e.g., in the areas where the virus normally replicates such as lungs, brain, heart, and kidney.

[00306] Methods for preventing a *coronavirus* infection prophylactically in a subject may comprise administering to the subject a prophylactically effective amount of a pan-coronavirus vaccine composition according to the present invention. In some embodiments, the composition elicits an immune response in the subject. In some embodiments, the composition induces memory B and T cells. In some embodiments, the composition induces resident memory T cells (Trm). In some embodiments, the composition prevents virus replication, e.g., in the areas where the virus normally replicates such as lungs, brain, heart, and kidney. In some embodiments, the composition prevents a cytokine storm, e.g., in the areas where the virus normally replicates such as lungs, brain, heart, and kidney. In some embodiments, the composition prevents inflammation or an inflammatory response, e.g., in the areas where the virus normally replicates such as lungs, brain, heart, and kidney. In some embodiments, the composition improves homing and retention of T cells, e.g., in the areas where the virus normally replicates such as lungs, brain, heart, and kidney.

[00307] Methods for eliciting an immune response in a subject may comprise administering to the subject

a vaccine composition according to the present invention, wherein the composition elicits an immune response in the subject. In some embodiments, the composition induces memory B and T cells. In some embodiments, the composition induces resident memory T cells (Trm). In some embodiments, the composition prevents virus replication, e.g., in the areas where the virus normally replicates such as lungs, brain, heart, and kidney. In some embodiments, the composition prevents a cytokine storm, e.g., in the areas where the virus normally replicates such as lungs, brain, heart, and kidney. In some embodiments, the composition prevents inflammation or an inflammatory response, e.g., in the areas where the virus normally replicates such as lungs, brain, heart, and kidney. In some embodiments, the composition improves homing and retention of T cells, e.g., in the areas where the virus normally replicates such as lungs, brain, heart, and kidney.

[00308] Methods for prolonging an immune response induced by a vaccine composition of the present invention and increasing T cell migration to particular tissues (e.g., lung, brain, heart, kidney, etc.) may comprise co-expressing a T-cell attracting chemokine, a composition that promotes T cell proliferation, and a vaccine composition (e.g., antigen) according to the present invention.

[00309] Methods for prolonging the retention of memory T-cell into the lungs induced by a vaccine composition of the present invention and increasing virus-specific tissue resident memory T-cells (T_{RM} cells) may comprise co-expressing a T-cell attracting chemokine, a composition that promotes T cell proliferation, and a vaccine composition (e.g., antigen) according to the present invention.

[00310] The vaccine composition may be administered through standard means, e.g., through an intravenous route (i.v.), an intranasal route (i.n.), or a sublingual route (s.l.) route.

[00311] In certain embodiments, the method comprises administering to the subject a second (e.g., booster) dose. The second dose may comprise the same vaccine composition or a different vaccine composition. Additional doses of one or more vaccine compositions may be administered.

Sequential Vaccine Delivery Methodology

[00312] In some embodiments, the present invention features a method of delivering the vaccine to induce heterologous immunity in a subject (e.g., prime/boost, see FIG. 25B and FIG. 26B). In some embodiments, the method comprises administering a first pan-coronavirus vaccine composition dose using a first delivery system. In further embodiments, the method comprises administering a second vaccine composition dose using a second delivery system. In some embodiments, the second composition is administered 8 days after administration of the first composition. In some embodiments, the second composition is administered 9 days after administration of the first composition. In some embodiments, the second composition is administered 10 days after administration of the first composition. In some embodiments, the second composition is administered 11 days after administration of the first composition. In some embodiments, the second composition is administered 12 days after administration of the first composition. In some embodiments, the second composition is administered 13 days after administration of the first composition. In some embodiments, the second composition is administered 14 days after administration of the first composition. In some embodiments, the second

composition is administered from 14 to 30 days after administration of the first composition. In some embodiments, the second composition is administered from 30 to 60 days after administration of the first composition. In other embodiments, the first delivery system and the second delivery system are different. In some embodiments, the peptide vaccine composition is administered 14-days after the administration of the first vaccine composition dose. In some embodiments, the peptide vaccine composition is administered 30 or 60 days after the administration of the first vaccine composition dose.

[00313] In some embodiments, the first delivery system or the second delivery system comprises an mRNA, a modified mRNA or a peptide vector. In other embodiments, the peptide vector comprises adenovirus or an adeno-associated virus vector.

[00314] In some embodiments, the present invention features a method of delivering the vaccine to induce heterologous immunity in a subject (i.e. prime/pull, see FIG. 25A and FIG. 26A). In some embodiments, the method comprises administering a pan-coronavirus vaccine composition. In further embodiments, the method comprises administering at least one T-cell attracting chemokine after administering the pan-coronavirus vaccine composition. In some embodiments, the T-cell attracting chemokine is administered 8 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 9 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 10 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 11 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 12 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 13 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 14 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered from 14 to 30 days after administration of the vaccine composition. In some embodiments, the T-cell attracting chemokine is administered from 30 to 60 days after administration of the vaccine composition. In some embodiments, the T cell-attracting chemokine composition is administered 8 to 14-days after the administration of the final vaccine composition dose. In some embodiments, the cell-attracting chemokine composition is administered 30 or 60 days after the administration of the final vaccine composition dose.

[00315] The present invention also features a novel "prime, pull, and boost" strategy. In other embodiments, the present invention features a method to increase the size and maintenance of lung-resident B-cells, CD4+ T cells and CD8+ T cells to protect against SARS-CoV-2 (FIG. 25D and FIG. 26D). In some embodiments, the method comprises administering a pan-coronavirus vaccine composition. In other embodiments, the method comprises administering at least one T-cell attracting chemokine after administering the pan-coronavirus vaccine composition. In further embodiments, the method comprises administering at least one cytokine after administering the T-cell attracting chemokine. In some embodiments, the T-cell attracting chemokine is administered 14 days after administering the pan-coronavirus composition. In other embodiments, the cytokine is administered 10 days after administering the T-cell attracting chemokine. In some embodiments, the T-cell attracting chemokine is

administered 8 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 9 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 10 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 11 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 12 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 13 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 14 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered from 14 to 30 days after administration of the vaccine composition. In some embodiments, the T-cell attracting chemokine is administered from 30 to 60 days after administration of the vaccine composition. In some embodiments, the cytokine is administered 8 days after administering the T-cell attracting chemokine. In some embodiments, the cytokine is administered 9 days after administering the T-cell attracting chemokine. In some embodiments, the cytokine is administered 10 days after administering the T-cell attracting chemokine. In some embodiments, the cytokine is administered 11 days after administering the T-cell attracting chemokine. In some embodiments, the cytokine is administered 12 days after administering the T-cell attracting chemokine. In some embodiments, the cytokine is administered 13 days after administering the T-cell attracting chemokine. In some embodiments, the cytokine is administered 14 days after administering the T-cell attracting chemokine. In some embodiments, the cytokine is administered from 14 to 30 days after administering the T-cell attracting chemokine. In some embodiments, the cytokine is administered from 30 to 60 days after administering the T-cell attracting chemokine. In some embodiments, the cytokine composition is administered 8 to 14-days after the administration of the T cell-attracting chemokine. In some embodiments, the cytokine composition is administered 30 or 60 days after the administration of the T cell-attracting chemokine.

[00316] The present invention further features a novel “prime, pull, and keep” strategy (FIG. 25C and FIG. 26C). In further embodiments, the present invention features a method to increase the size and maintenance of lung-resident B-cells, CD4+ T cells and CD8+ T cells to protect against SARS-CoV-2. In some embodiments, the method comprises administering a pan-*coronavirus* vaccine composition. In other embodiments, the method comprises administering at least one T-cell attracting chemokine after administering the pan-*coronavirus* vaccine composition. In further embodiments, the method comprises administering at least one mucosal chemokine after administering the T-cell attracting chemokine. In some embodiments, the T-cell attracting chemokine is administered 14 days after administering the pan-*coronavirus* composition. In other embodiments, the mucosal chemokines is administered 10 days after administering the T-cell attracting chemokine. In some embodiments, the T-cell attracting chemokine is administered 8 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 9 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 10 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 11 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is

administered 12 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 13 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered 14 days after the vaccine composition is administered. In some embodiments, the T-cell attracting chemokine is administered from 14 to 30 days after administration of the vaccine composition. In some embodiments, the T-cell attracting chemokine is administered from 30 to 60 days after administration of the vaccine composition. In some embodiments, the mucosal chemokine is administered 8 days after administering the T-cell attracting chemokine. In some embodiments, the mucosal chemokine is administered 9 days after administering the T-cell attracting chemokine. In some embodiments, the mucosal chemokine is administered 10 days after administering the T-cell attracting chemokine. In some embodiments, the mucosal chemokine is administered 11 days after administering the T-cell attracting chemokine. In some embodiments, the mucosal chemokine is administered 12 days after administering the T-cell attracting chemokine. In some embodiments, the mucosal chemokine is administered 13 days after administering the T-cell attracting chemokine. In some embodiments, the mucosal chemokine is administered 14 days after administering the T-cell attracting chemokine. In some embodiments, the mucosal chemokine is administered from 14 to 30 days after administering the T-cell attracting chemokine. In some embodiments, the mucosal chemokine is administered from 30 to 60 days after administering the T-cell attracting chemokine. In some embodiments, the mucosal chemokine composition is administered 8 to 14-days after the administration of the T cell-attracting chemokine. In some embodiments, the mucosal cytokine composition is administered 30 or 60 days after the administration of the T cell-attracting chemokine.

[00317] In some embodiments, the mucosal chemokines may comprise CCL25, CCL28, CXCL14, CXCL17, or a combination thereof. In some embodiments, the T-cell attracting chemokines may comprise CCL5, CXCL9, CXCL10, CXCL11, or a combination thereof. In some embodiments, the cytokines may comprise IL-15, IL-2, IL-7 or a combination thereof.

[00318] In some embodiments, the efficacy (or effectiveness) of a vaccine composition herein is greater than 60%. In some embodiments, the efficacy (or effectiveness) of a vaccine composition herein is greater than 70%. In some embodiments, the efficacy (or effectiveness) of a vaccine composition herein is greater than 80%. In some embodiments, the efficacy (or effectiveness) of a vaccine composition herein is greater than 90%. In some embodiments, the efficacy (or effectiveness) of a vaccine composition herein is greater than 95%.

[00319] Vaccine efficacy may be assessed using standard analyses (see, e.g., Weinberg et al., J Infect Dis. 2010 Jun. 1; 201(11):1607-10). For example, vaccine efficacy may be measured by double-blind, randomized, clinical controlled trials. Vaccine efficacy may be expressed as a proportionate reduction in disease attack rate (AR) between the unvaccinated (ARU) and vaccinated (ARV) study cohorts and can be calculated from the relative risk (RR) of disease among the vaccinated group with use of the following formulas: $\text{Efficacy} = (\text{ARU} - \text{ARV}) / \text{ARU} \times 100$; and $\text{Efficacy} = (1 - \text{RR}) \times 100$.

[00320] Likewise, vaccine effectiveness may be assessed using standard analyses (see, e.g., Weinberg

et al., J Infect Dis. 2010 Jun. 1; 201(11):1607-10). Vaccine effectiveness is an assessment of how a vaccine (which may have already proven to have high vaccine efficacy) reduces disease in a population. This measure can assess the net balance of benefits and adverse effects of a vaccination program, not just the vaccine itself, under natural field conditions rather than in a controlled clinical trial. Vaccine effectiveness is proportional to vaccine efficacy (potency) but is also affected by how well target groups in the population are immunized, as well as by other non-vaccine-related factors that influence the 'real-world' outcomes of hospitalizations, ambulatory visits, or costs. For example, a retrospective case control analysis may be used, in which the rates of vaccination among a set of infected cases and appropriate controls are compared. Vaccine effectiveness may be expressed as a rate difference, with use of the odds ratio (OR) for developing infection despite vaccination: $\text{Effectiveness} = (1 - \text{OR}) \times 100$.

[00321] In some embodiments, the vaccine immunizes the subject against a coronavirus for up to 1 year. In some embodiments, the vaccine immunizes the subject against a coronavirus for up to 2 years. In some embodiments, the vaccine immunizes the subject against a coronavirus for more than 1 year, more than 2 years, more than 3 years, more than 4 years, or for 5-10 years.

[00322] In some embodiments, the subject is a young adult between the ages of about 20 years and about 50 years (e.g., about 20, 25, 30, 35, 40, 45 or 50 years old).

[00323] In some embodiments, the subject is an elderly subject about 60 years old, about 70 years old, or older (e.g., about 60, 65, 70, 75, 80, 85 or 90 years old).

[00324] In some embodiments, the subject is about 5 years old or younger. For example, the subject may be between the ages of about 1 year and about 5 years (e.g., about 1, 2, 3, 4 or 5 years), or between the ages of about 6 months and about 1 year (e.g., about 6, 7, 8, 9, 10, 11 or 12 months). In some embodiments, the subject is about 12 months or younger (e.g., 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 months or 1 month). In some embodiments, the subject is about 6 months or younger.

[00325] In some embodiments, the subject was born full term (e.g., about 37-42 weeks). In some embodiments, the subject was born prematurely, for example, at about 36 weeks of gestation or earlier (e.g., about 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26 or 25 weeks). For example, the subject may have been born at about 32 weeks of gestation or earlier. In some embodiments, the subject was born prematurely between about 32 weeks and about 36 weeks of gestation. In such subjects, a vaccine may be administered later in life, for example, at the age of about 6 months to about 5 years, or older.

[00326] In some embodiments, the subject is pregnant (e.g., in the first, second or third trimester) when administered a vaccine.

[00327] In some embodiments, the subject has a chronic pulmonary disease (e.g., chronic obstructive pulmonary disease (COPD) or asthma) or is at risk thereof. Two forms of COPD include chronic bronchitis, which involves a long-term cough with mucus, and emphysema, which involves damage to the lungs over time. Thus, a subject administered a vaccine may have chronic bronchitis or emphysema.

[00328] In some embodiments, the subject has been exposed to a coronavirus.. In some embodiments, the subject is infected with a coronavirus. In some embodiments, the subject is at risk of infection by a coronavirus.

[00329] In some embodiments, the subject is immunocompromised (has an impaired immune system, e.g., has an immune disorder or autoimmune disorder).

Pharmaceutical Carriers

[00330] In certain embodiments, the vaccine composition further comprises a pharmaceutical carrier. Pharmaceutical carriers are well known to one of ordinary skill in the art. For example, in certain embodiments, the pharmaceutical carrier is selected from the group consisting of water, an alcohol, a natural or hardened oil, a natural or hardened wax, a calcium carbonate, a sodium carbonate, a calcium phosphate, kaolin, talc, lactose and combinations thereof. In some embodiments, the pharmaceutical carrier may comprise a lipid nanoparticle, an adenovirus vector, or an adeno-associated virus vector. In some embodiments, the vaccine composition is constructed using an adeno-associated virus vectors-based antigen delivery system.

[00331] Also provided herein is vaccine of any one of the foregoing paragraphs, formulated in a nanoparticle (e.g., a lipid nanoparticle). In some embodiments, the nanoparticle has a mean diameter of 50-200 nm. In some embodiments, the nanoparticle is a lipid nanoparticle. In some embodiments, the lipid nanoparticle comprises a cationic lipid, a PEG-modified lipid, a sterol and a non-cationic lipid. In some embodiments, the lipid nanoparticle comprises a molar ratio of about 20-60% cationic lipid, 0.5-15% PEG-modified lipid, 25-55% sterol, and 25% non-cationic lipid. In some embodiments, the cationic lipid is an ionizable cationic lipid and the non-cationic lipid is a neutral lipid, and the sterol is a cholesterol. In some embodiments, the cationic lipid is selected from 2,2-dilinoyleyl-4-dimethylaminoethyl-[1,3]-dioxolane (DLin-KC2-DMA), dilinoyleyl-methyl-4-dimethylaminobutyrate (DLin-MC3-DMA), and di((Z)-non-2-en-1-yl) 9-((4-(dimethylamino)butanoyl)oxy)heptadecanedioate (L319).

[00332] Although there has been shown and described the preferred embodiment of the present invention, it will be readily apparent to those skilled in the art that modifications may be made thereto which do not exceed the scope of the appended claims. Therefore, the scope of the invention is only to be limited by the following claims. In some embodiments, the figures presented in this patent application are drawn to scale, including the angles, ratios of dimensions, etc. In some embodiments, the figures are representative only and the claims are not limited by the dimensions of the figures. In some embodiments, descriptions of the inventions described herein using the phrase "comprising" includes embodiments that could be described as "consisting essentially of" or "consisting of", and as such the written description requirement for claiming one or more embodiments of the present invention using the phrase "consisting essentially of" or "consisting of" is met.

WHAT IS CLAIMED:

1. A pan-*coronavirus* recombinant vaccine composition, the composition comprising one or more large sequences, wherein each of the one or more large sequences comprise at least one of:
 - a) one or more conserved *coronavirus* B-cell target epitopes;
 - b) one or more conserved *coronavirus* CD4⁺ T cell target epitopes; and/or
 - c) one or more conserved *coronavirus* CD8⁺ T cell target epitopes;wherein at least one epitope is derived from a non-spike protein.
2. A pan-*coronavirus* recombinant vaccine composition, the composition comprising two or more large sequences, wherein each of the two or more large sequences comprise at least one of:
 - a) one or more conserved *coronavirus* B-cell target epitopes;
 - b) one or more conserved *coronavirus* CD4⁺ T cell target epitopes; and/or
 - c) one or more conserved *coronavirus* CD8⁺ T cell target epitopes;wherein at least one epitope is derived from a non-spike protein.
3. A pan-*coronavirus* recombinant vaccine composition, the composition comprising whole spike protein; and one or both of:
 - a) one or more conserved *coronavirus* CD4⁺ T cell target epitopes; and/or
 - b) one or more conserved *coronavirus* CD8⁺ T cell target epitopes;wherein at least one epitope is derived from a non-spike protein.
4. A pan-*coronavirus* recombinant vaccine composition, the composition comprising at least a portion of spike protein, the portion of spike protein comprising a trimerized SARS-CoV-2 receptor-binding domain (RBD); and one or both of:
 - (a) one or more conserved *coronavirus* CD4⁺ T cell target epitopes;
 - (b) one or more conserved *coronavirus* CD8⁺ T cell target epitopes;wherein at least one epitope is derived from a non-spike protein.
5. A pan-*coronavirus* recombinant vaccine composition, the composition comprising whole spike protein; and one or more conserved *coronavirus* CD4⁺ T cell target epitopes; and one or more conserved *coronavirus* CD8⁺ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.
6. A pan-*coronavirus* recombinant vaccine composition, the composition comprising at least a portion of spike protein, the portion of spike protein comprising a trimerized SARS-CoV-2 receptor-binding domain (RBD); and one or more conserved *coronavirus* CD4⁺ T cell target epitopes; and one or more conserved *coronavirus* CD8⁺ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.
7. The composition of any of claims 1-6, wherein the non-spike protein is ORF1ab protein, ORF3a protein, Envelope protein, Membrane glycoprotein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, Nucleocapsid protein and ORF10 protein.
8. The composition of any of claims 1-7, wherein the one or more large sequences are highly conserved among human and animal *coronaviruses*.
9. The composition of any of claims 1-8, wherein the one or more large sequences are derived from

at least one of SARS-CoV-2 protein.

10. The composition of any of claims 1-9, wherein the one or more large sequences are derived from one or more of: one or more SARS-CoV-2 human strains or variants in current circulation; one or more *coronaviruses* that has caused a previous human outbreak; one or more *coronaviruses* isolated from animals selected from a group consisting of bats, pangolins, civet cats, minks, camels, and other animal receptive to coronaviruses; or one or more *coronaviruses* that cause the common cold.
11. The composition of claim 10, wherein the one or more SARS-CoV-2 human strains or variants in current circulation are selected from: strain B.1.177; strain B.1.160, strain B.1.1.7; strain B.1.351; strain P.1; strain B.1.427/B.1.429; strain B.1.258; strain B.1.221; strain B.1.367; strain B.1.1.277; strain B.1.1.302; strain B.1.525; strain B.1.526, strain S:677H, and strain S:677P.
12. The composition of claim 10, wherein the one or more *coronaviruses* that cause the common cold are selected from: 229E alpha coronavirus, NL63 alpha coronavirus, OC43 beta coronavirus, and HKU1 beta coronavirus.
13. The composition of any of claims 1-12, wherein the conserved epitopes are selected from Variants Of Concern or Variants Of Interest.
14. The composition of any of claims 1-13, wherein the one or more large sequences are derived from a whole protein sequence expressed by SARS-CoV-2.
15. The composition of any of claims 1-14, wherein the one or more large sequences are derived from a partial protein sequence expressed by SARS-CoV-2.
16. The composition of any of claims 1-15, wherein the one or more large conserved sequences is derived from a full-length spike glycoprotein.
17. The composition of any of claims 1-15, wherein the one or more large conserved sequences is derived from a partial spike glycoprotein.
18. The composition of claim 16-17, wherein the spike protein has two consecutive proline substitutions at amino acid positions 986 and 987.
19. The composition of any of claims 16-18, wherein the spike glycoprotein has single amino acid substitutions at amino acid positions comprising Tyr-83 and Tyr-489, Gln-24 and Asn-487.
20. The composition of any of claims 1-19, wherein the one or more large sequences comprises Spike glycoprotein (S) or a portion thereof, Nucleoprotein or a portion thereof, Membrane protein or a portion thereof, and ORF1a/b or a portion thereof.
21. The composition of any of claims 1-19, wherein the one or more large sequences comprises Spike glycoprotein (S) or a portion thereof, Nucleoprotein or a portion thereof, and ORF1a/b or a portion thereof.
22. The composition of any of claims 16-19, wherein the portion of the Spike glycoprotein is RBD.
23. The composition of any of claims 1-22, wherein the one or more large sequences is selected from the group consisting of: ORF1ab protein, Spike glycoprotein, ORF3a protein, Envelope protein, Membrane glycoprotein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, Nucleocapsid protein and ORF10 protein.
24. The composition of claim 23 wherein the ORF1ab protein comprises nonstructural protein (Nsp)

- 1, Nsp2, Nsp3, Nsp4, Nsp5, Nsp6, Nsp7, Nsp8, Nsp9, Nsp10, Nsp11, Nsp12, Nsp13, Nsp14, Nsp15 and Nsp16.
25. The composition of any of claims 1-24, wherein one or more of the large sequences comprises a T-cell epitope restricted to a large number of human class 1 and class 2 HLA haplotypes and are not restricted to HLA-0201 for class 1 or HLA-DR for class 2.
26. The composition of any of claims 1-25, wherein the large sequences are derived from structural proteins, non-structural proteins, or a combination thereof.
27. The composition of any of claims 1-26, wherein one or more of the large sequences are derived from a protein selected from SEQ ID NO: 182-185, SEQ ID NO: 148-159, SEQ ID NO: 186-187, and SEQ ID NO: 191-196.
28. The composition of any of claims 1-27, wherein the one or more conserved *coronavirus* CD8⁺ T cell target epitopes are selected from: spike glycoprotein, Envelope protein, ORF1ab protein, ORF7a protein, ORF8a protein, ORF10 protein, or a combination thereof.
29. The composition of any of claims 1-28, wherein the one or more conserved *coronavirus* CD8⁺ T cell target epitopes are selected from: S₂₋₁₀, S₁₂₂₀₋₁₂₂₈, S₁₀₀₀₋₁₀₀₈, S₉₅₈₋₉₆₆, E₂₀₋₂₈, ORF1ab₁₆₇₅₋₁₆₈₃, ORF1ab₂₃₆₃₋₂₃₇₁, ORF1ab₃₀₁₃₋₃₀₂₁, ORF1ab₃₁₈₃₋₃₁₉₁, ORF1ab₅₄₇₀₋₅₄₇₈, ORF1ab₆₇₄₉₋₆₇₅₇, ORF7b₂₆₋₃₄, ORF8a₇₃₋₈₁, ORF10₃₋₁₁, and ORF10₅₋₁₃.
30. The composition of any of claims 1-29, wherein the one or more conserved *coronavirus* CD8⁺ T cell target epitopes are selected from SEQ ID NO: 2-29.
31. The composition of any of claims 1-29, wherein the one or more conserved *coronavirus* CD8⁺ T cell target epitopes are selected from SEQ ID NO: 30-57.
32. The composition of any of claims 1-31, wherein the one or more conserved *coronavirus* CD4⁺ T cell target epitopes are selected from: spike glycoprotein, Envelope protein, Membrane protein, Nucleocapsid protein, ORF1a protein, ORF1ab protein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, or a combination thereof.
33. The composition of any of claims 1-32, wherein the one or more conserved *coronavirus* CD4⁺ T cell target epitopes are selected from: ORF1a₁₃₅₀₋₁₃₆₅, ORF1ab₅₀₁₉₋₅₀₃₃, ORF6₁₂₋₂₆, ORF1ab₆₀₈₈₋₆₁₀₂, ORF1ab₆₄₂₀₋₆₄₃₄, ORF1a₁₈₀₁₋₁₈₁₅, S₁₋₁₃, E₂₆₋₄₀, E₂₀₋₃₄, M₁₇₆₋₁₉₀, N₃₈₈₋₄₀₃, ORF7a₃₋₁₇, ORF7a₁₋₁₅, ORF7b₈₋₂₂, ORF7a₉₈₋₁₁₂, and ORF8₁₋₁₅.
34. The composition of any of claims 1-33, wherein the one or more conserved *coronavirus* CD4⁺ T cell target epitopes are selected from SEQ ID NO: 58-73.
35. The composition of any of claims 1-33, wherein the one or more conserved *coronavirus* CD4⁺ T cell target epitopes are selected from SEQ ID NO: 74-105.
36. The composition of any of claims 1-35, wherein the one or more conserved *coronavirus* B cell target epitopes are selected from Spike glycoprotein.
37. The composition of any of claims 1-36, wherein the one or more conserved *coronavirus* B cell target epitopes are selected from: S₂₈₇₋₃₁₇, S₅₂₄₋₅₉₈, S₆₀₁₋₆₄₀, S₈₀₂₋₈₁₉, S₈₈₈₋₉₀₉, S₃₆₉₋₃₉₃, S₄₄₀₋₅₀₁, S₁₁₃₃₋₁₁₇₂, S₃₂₉₋₃₆₃, and S₁₃₋₃₇.
38. The composition of any of claims 1-37, wherein the one or more *coronavirus* B cell target epitopes are selected from SEQ ID NO: 106-116.

39. The composition of any of claims 1-37, wherein the one or more *coronavirus* B cell target epitopes are selected from SEQ ID NO: 117-138.
40. The composition of any of claims 1-39, wherein the one or more conserved coronavirus B cell target epitopes are in the form of a large sequence.
41. The composition of claim 40, wherein the large sequence is full length spike glycoprotein.
42. The composition of claim 40, wherein the large sequence is a partial spike glycoprotein.
43. The composition of any of claims 1-42, wherein each of the large sequences are separated by a linker.
44. The composition of claim 43, wherein the linker is from 2 to 10 amino acids in length.
45. The composition of any of claims 1-44 further comprising a T cell attracting chemokine.
46. The composition of claim 45, wherein the T cell attracting chemokine is CCL5, CXCL9, CXCL10, CXCL11, or a combination thereof.
47. The composition of any of claims 1-46 further comprising a composition that promotes T cell proliferation.
48. The composition of claim 47, wherein the composition that promotes T cell proliferation is IL-7, IL-2, or IL-15.
49. The composition of any of claims 1-49 further comprising a molecular adjuvant.
50. The composition of claim 50, wherein the molecular adjuvant is CpG.
51. The composition of claim 54, wherein the molecular adjuvant is a CpG polymer.
52. The composition of claim 50, wherein the molecular adjuvant is flagellin.
53. The composition of any of claims 1-52, wherein the recombinant vaccine composition comprises a tag.
54. The composition of claim 53, wherein the tag is a His tag.
55. The composition of any of claims 1-54, further comprising a pharmaceutical carrier.
56. The composition of any of claims 1-55, wherein the composition is used to prevent a *coronavirus* disease in a subject.
57. The composition of any of claims 1-56, where the composition is used to prevent a *coronavirus* infection prophylactically in a subject.
58. The composition of any of claims 1-57, wherein the composition elicits an immune response in a subject.
59. The composition of any of claims 1-58, wherein the composition prolongs an immune response induced by the multi-epitope pan-*coronavirus* recombinant vaccine composition and increases T-cell migration to the lungs.
60. The composition of any of claims 1-59, wherein the transmembrane anchor of the spike protein has an intact S1-S2 cleavage site.
61. The composition of any of claims 1-60, wherein the spike protein is in its stabilized conformation.
62. The composition of any of claims 1-61, wherein the spike protein is stabilized with proline substitutions at amino acid positions 986 and 987 at the top of the central helix in the S2 subunit.
63. A recombinant vaccine composition according to any of claims 1-62 comprising full-length spike protein.

64. A recombinant vaccine composition according to any of claims 1-62 comprising full-length spike protein or partial spike protein.
65. The composition of any of claims 1-64, wherein the spike protein comprises Tyr-489 and Asn-487.
66. The composition of any of claims 1-65, wherein the spike protein comprises Gln-493.
67. The composition of any of claims 1-66, wherein the spike protein comprises Tyr-505.
68. The composition of any of claims 1-67, comprising a trimerized SARS-CoV-2 receptor-binding domain (RBD) sequence.
69. The composition of claim 68, wherein the trimerized SARS-CoV-2 receptor-binding domain (RBD) sequence is modified by the addition of a T4 fibrin-derived foldon trimerization domain.
70. The composition of claim 69, wherein the addition of a T4 fibrin-derived foldon trimerization domain increases immunogenicity by multivalent display.
71. The composition of any of claims 1-70, wherein the conserved epitopes are selected from the Variants Of Concern and Variants Of Interest.
72. The composition of any of claims 1-71 comprising a mutation 682-RRAR-685 → 682-QQAQ-685 in the S1-S2 cleavage site.
73. The composition of any of claims 1-72 comprising at least one proline substitution.
74. The composition of any of claims 1-72 comprising at least two proline substitutions.
75. The composition of any of claims 73-74, wherein the proline substitution is at position K986 and V987.
76. The composition of any of claims 1-75 comprising K986P and V987P mutations.
77. The composition of any of claims 1-83, wherein the spike protein or a portion thereof comprises SEQ ID NO: 195.
78. The composition of any of claims 1-83, wherein the spike protein or a portion thereof comprises one of SEQ ID NO: 191 or SEQ ID NO: 192.
79. The composition of any of claims 1-83, wherein the spike protein or a portion thereof comprises one of SEQ ID NO: 186-187 and SEQ ID NO: 191-196.
80. A pan-coronavirus recombinant vaccine composition, the composition comprising an antigen delivery system encoding one or more large sequences, wherein each of the one or more large sequences comprise at least one of:
 - a) one or more conserved coronavirus B-cell target epitopes;
 - b) one or more conserved coronavirus CD4⁺ T cell target epitopes; and/or
 - c) one or more conserved coronavirus CD8⁺ T cell target epitopes;
 wherein at least one epitope is from a non-spike protein.
81. A pan-coronavirus recombinant vaccine composition, the composition comprising an antigen delivery system encoding two or more large sequences, wherein each of the two or more large sequences comprise at least one of:
 - a) one or more conserved coronavirus B-cell target epitopes;
 - b) one or more conserved coronavirus CD4⁺ T cell target epitopes; and/or
 - c) one or more conserved coronavirus CD8⁺ T cell target epitopes;

- wherein at least one epitope is derived from a non-spike protein.
82. A pan-coronavirus recombinant vaccine composition, the composition comprising an antigen delivery system encoding whole spike protein; and one or both of:
- a) one or more conserved coronavirus CD4⁺ T cell target epitopes; and/or
 - b) one or more conserved coronavirus CD8⁺ T cell target epitopes;
- wherein at least one epitope is derived from a non-spike protein.
83. A pan-coronavirus recombinant vaccine composition, the composition comprising an antigen delivery system encoding at least a portion of spike protein, the portion of spike protein comprising a trimerized SARS-CoV-2 receptor-binding domain (RBD); and one or both of:
- (c) one or more conserved coronavirus CD4⁺ T cell target epitopes;
 - (d) one or more conserved coronavirus CD8⁺ T cell target epitopes;
- wherein at least one epitope is derived from a non-spike protein.
84. A pan-coronavirus recombinant vaccine composition, the composition comprising an antigen delivery system encoding whole spike protein; and one or more conserved coronavirus CD4⁺ T cell target epitopes; and one or more conserved coronavirus CD8⁺ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.
85. A pan-coronavirus recombinant vaccine composition, the composition comprising an antigen delivery system encoding at least a portion of spike protein, the portion of spike protein comprising a trimerized SARS-CoV-2 receptor-binding domain (RBD); and one or more conserved coronavirus CD4⁺ T cell target epitopes; and one or more conserved coronavirus CD8⁺ T cell target epitopes; wherein at least one epitope is derived from a non-spike protein.
86. The composition of any of claims 80-85, wherein the composition comprises two or more large sequences.
87. The composition of any of claims 80-86, wherein the delivery system is an adenovirus system.
88. The composition of claim 87, wherein the adenovirus delivery system is Ad26, Ad5, Ad35, or a combination thereof.
89. The composition of any of claims 80-88, wherein the one or more large sequences are operatively linked to a generic promoter.
90. The composition of claim 89, wherein the generic promoter is a CMV or a CAG promoter.
91. The composition of any of claims 80-88, wherein the one or more large sequences are operatively linked to a lung-specific promoter.
92. The composition of claim 91, wherein the lung-specific promoter is SpB or CD144.
93. The composition of any of claims 80-92, wherein the antigen delivery system further encodes a T cell attracting chemokine.
94. The composition of claim 93, wherein the T cell attracting chemokine is CCL5, CXCL9, CXCL10, CXCL11, or a combination thereof.
95. The composition of any of claims 93-94, wherein the T cell attracting chemokine is operatively linked to a lung-specific promoter.
96. The composition of any of claims 93-94, wherein the T cell attracting chemokine is operatively linked to a generic promoter.

97. The composition of any of claims 80-96, wherein the antigen delivery system further encodes a composition that promotes T cell proliferation.
98. The composition of claim 97, wherein the composition that promotes T cell proliferation is IL-7, IL-2, or IL-15.
99. The composition of any of claims 97-98, wherein the composition that promotes T cell proliferation is operatively linked to a lung-specific promoter.
100. The composition of any of claims 97-98, wherein the composition that promotes T cell proliferation is operatively linked to a generic promoter.
101. The composition of any of claims 93-100, wherein the T cell attracting chemokine and the composition that promotes T cell proliferation are driven by the same promoter.
102. The composition of any of claims 80-101, wherein the vaccine further encodes a peptide comprising a T cell attracting chemokine and a composition that promotes T cell proliferation.
103. The composition of claim 102, wherein the peptide is operatively linked to a lung-specific promoter.
104. The composition of claim 102 wherein the peptide is operatively linked to a generic promoter.
105. The composition of any of claims 80-104, wherein the antigen delivery system further encodes a molecular adjuvant.
106. The composition of claim 105, wherein the molecular adjuvant is CpG.
107. The composition of claim 106, wherein the molecular adjuvant is a CpG polymer.
108. The composition of claim 105, wherein the molecular adjuvant is flagellin.
109. The composition of any of claims 105-108, wherein the molecular adjuvant is operatively linked to a promoter.
110. The composition of claim 109, wherein the promoter is a lung-specific promoter or a generic promoter.
111. The composition of any of claims 80-110, wherein the one or more large sequences are highly conserved among human and animal coronaviruses.
112. The composition of any of claims 80-111, wherein the one or more large sequences are derived from at least one of SARS-CoV-2 protein.
113. The composition of any of claims 80-112, wherein the one or more large sequences are derived from one or more of: (a) one or more SARS-CoV-2 human strains or variants in current circulation; (b) one or more *coronaviruses* that has caused a previous human outbreak; (c) one or more *coronaviruses* isolated from animals selected from a group consisting of bats, pangolins, civet cats, minks, camels, and other animal receptive to coronaviruses; or (d) one or more *coronaviruses* that cause the common cold.
114. The composition of claim 113, wherein the one or more SARS-CoV-2 human strains or variants in current circulation are selected from: strain B.1.177; strain B.1.160, strain B.1.1.7; strain B.1.351; strain P.1; strain B.1.427/B.1.429; strain B.1.258; strain B.1.221; strain B.1.367; strain B.1.1.277; strain B.1.1.302; strain B.1.525; strain B.1.526, strain S:677H, and strain S:677P.
115. The composition of claim 113, wherein the one or more *coronaviruses* that cause the common

could be selected from: 229E alpha coronavirus, NL63 alpha coronavirus, OC43 beta coronavirus, and HKU1 beta coronavirus.

116. The composition of any of claims 80-115, wherein each of the large sequences are separated by a linker.
117. The composition of claim 116, wherein the linker is from 2 to 10 amino acids in length.
118. The composition of any of claims 80-117, wherein the recombinant vaccine composition comprises a tag.
119. The composition of claim 118, wherein the tag is a His tag.
120. The composition of any of claims 80-119, wherein the large sequences are derived from structural proteins, non-structural proteins, or a combination thereof.
121. The composition of any of claims 80-120, wherein target epitopes are derived from a SARS-CoV-2 protein selected from a group consisting of: ORF1ab protein, Spike glycoprotein, ORF3a protein, Envelope protein, Membrane glycoprotein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, Nucleocapsid protein and ORF10 protein.
122. The composition of claim 121, wherein the target epitope derived from the Spike glycoprotein is RBD.
123. The composition of claim 121, wherein the target epitope derived from the Spike glycoprotein is NTD.
124. The composition of claim 121, wherein the target epitope derived from the Spike glycoprotein includes both the RBD and NTD regions.
125. The composition of any of claims 121-124, wherein the target epitope derived from the spike glycoprotein are recognized by neutralizing and blocking antibodies.
126. The composition of any of claims 121-124, wherein the target epitope derived from the spike glycoprotein induces neutralizing and blocking antibodies.
127. The composition of any of claims 121-124, wherein the target epitope derived from the spike glycoprotein induces neutralizing and blocking antibodies that recognize and neutralize the virus.
128. The composition of any of claims 121-124, wherein the target epitope derived from the spike glycoprotein induces neutralizing and blocking antibodies that recognize the spike protein.
129. The composition of claim 121, wherein the ORF1ab protein comprises nonstructural protein (Nsp) 1, Nsp2, Nsp3, Nsp4, Nsp5, Nsp6, Nsp7, Nsp8, Nsp9, Nsp10, Nsp11, Nsp12, Nsp13, Nsp14, Nsp15 and Nsp16.
130. The composition of any of claims 80-129, wherein the one or more conserved *coronavirus* CD8⁺ T cell target epitopes are selected from: spike glycoprotein, Envelope protein, ORF1ab protein, ORF7a protein, ORF8a protein, ORF10 protein, or a combination thereof.
131. The composition of any of claims 80-130, wherein the one or more conserved *coronavirus* CD8⁺ T cell target epitopes are selected from: S₂₋₁₀, S₁₂₂₀₋₁₂₂₈, S₁₀₀₀₋₁₀₀₈, S₉₅₈₋₉₆₆, E₂₀₋₂₈, ORF1ab₁₆₇₅₋₁₆₈₃, ORF1ab₂₃₆₃₋₂₃₇₁, ORF1ab₃₀₁₃₋₃₀₂₁, ORF1ab₃₁₈₃₋₃₁₉₁, ORF1ab₅₄₇₀₋₅₄₇₈, ORF1ab₆₇₄₉₋₆₇₅₇, ORF7b₂₆₋₃₄, ORF8a₇₃₋₈₁, ORF10₃₋₁₁, and ORF10₅₋₁₃.
132. The composition of any of claims 80-131, wherein the one or more conserved *coronavirus* CD8⁺ T cell target epitopes are selected from SEQ ID NO: 2-29.

133. The composition of any of claims 80-131, wherein the one or more conserved *coronavirus* CD8⁺ T cell target epitopes are selected from SEQ ID NO: 30-57.
134. The composition of any of claims 80-133, wherein the one or more conserved *coronavirus* CD4⁺ T cell target epitopes are selected from: spike glycoprotein, Envelope protein, Membrane protein, Nucleocapsid protein, ORF1a protein, ORF1ab protein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, or a combination thereof.
135. The composition of any of claims 80-134, wherein the one or more conserved *coronavirus* CD4⁺ T cell target epitopes are selected from: ORF1a₁₃₅₀₋₁₃₆₅, ORF1ab₅₀₁₉₋₅₀₃₃, ORF6₁₂₋₂₆, ORF1ab₆₀₈₈₋₆₁₀₂, ORF1ab₆₄₂₀₋₆₄₃₄, ORF1a₁₈₀₁₋₁₈₁₅, S₁₋₁₃, E₂₆₋₄₀, E₂₀₋₃₄, M₁₇₆₋₁₉₀, N₃₈₈₋₄₀₃, ORF7a₉₋₁₇, ORF7a₁₋₁₅, ORF7b₈₋₂₂, ORF7a₉₈₋₁₁₂, and ORF8₁₋₁₅.
136. The composition of any of claims 80-135, wherein the one or more conserved *coronavirus* CD4⁺ T cell target epitopes are selected from SEQ ID NO: 58-73.
137. The composition of any of claims 80-135, wherein the one or more conserved *coronavirus* CD4⁺ T cell target epitopes are selected from SEQ ID NO: 74-105.
138. The composition of any of claims 80-136, wherein the one or more conserved coronavirus B cell target epitopes are selected from Spike glycoprotein.
139. The composition of any of claims 80-137, wherein the one or more conserved coronavirus B cell target epitopes are selected from: S₂₈₇₋₃₁₇, S₅₂₄₋₅₉₈, S₆₀₁₋₆₄₀, S₈₀₂₋₈₁₉, S₈₈₈₋₉₀₉, S₃₆₉₋₃₉₃, S₄₄₀₋₅₀₁, S₁₁₃₃₋₁₁₇₂, S₃₂₉₋₃₆₃, and S₁₃₋₃₇.
140. The composition of any of claims 80-138, wherein the one or more *coronavirus* B cell target epitopes are selected from SEQ ID NO: 106-116.
141. The composition of any of claims 80-138, wherein the one or more *coronavirus* B cell target epitopes are selected from SEQ ID NO: 117-138.
142. The composition of any of claims 80-141, wherein each of the large sequences are separated by a linker.
143. The composition of any of claims 80-141, wherein two or more of the large sequences are separated by a linker.
144. The composition of claim 142-143, wherein the linker is from 2 to 10 amino acids in length.
145. The composition of any of claims 142-144, wherein the linker comprises T2A.
146. The composition of any of claims 142-144, wherein the linker is selected from T2A, E2A, and P2A.
147. The composition of any of claims 80-146, wherein a different linker is disposed between each open reading frame.
148. The composition of any of claims 80-147, wherein the recombinant vaccine composition comprises a tag.
149. The composition of claim 148, wherein the tag is a His tag.
150. The composition of any of claims 80-149, further comprising a pharmaceutical carrier.
151. The composition of any of claims 80-150, wherein the composition is used to prevent a *coronavirus* disease in a subject.
152. The composition of any of claims 80-151, where the composition is used to prevent a

- coronavirus* infection prophylactically in a subject.
153. The composition of any of claims 80-152, wherein the composition elicits an immune response in a subject.
154. The composition of any of claims 80-153, wherein the composition prolongs an immune response induced by the multi-epitope pan-*coronavirus* recombinant vaccine composition and increases T-cell migration to the lungs.
155. The composition of any of claims 80-154, wherein the vaccine constructs are for humans.
156. The composition of claim 155, wherein the recombinant vaccine composition comprises human CXCL-11 and IL-7 or IL-2 or IL-15.
157. The compositions of any of claims 1-154, wherein the vaccine constructs are for animals.
158. The composition of claim 60, wherein the recombinant vaccine composition comprises animal CXCL-11 and IL-7 or IL-2 or IL-15.
159. The composition of claim 157-158, wherein the animals are cats and dogs.
160. The composition of any of claims 1-159, wherein the compositions are for use as a vaccine.
161. The composition of any of claims 1-160, wherein the compositions are for use as immunotherapy for the prevention and treatment of Coronaviruses infections and diseases.
162. A rVSV-panCoV recombinant vaccine composition according to any of claims 1-161.
163. A rAdV-panCoV recombinant vaccine composition according to any of claims 1-161.
164. A pan-*coronavirus* recombinant vaccine composition comprising any of SEQ ID NO: 139-147.
165. A method of preventing a *coronavirus* disease in a subject; the method comprising: administering to the subject a therapeutically effective amount of a pan-*coronavirus* recombinant vaccine composition according to any of claims 1-164, wherein the composition elicits an immune response in the subject.
166. A method of preventing a *coronavirus* infection prophylactically in a subject, the method comprising: administering to the subject a prophylactically effective amount of a pan-*coronavirus* recombinant vaccine composition according to any of claims 1-164.
167. A method of eliciting an immune response in a subject, comprising administering to the subject a composition according to any of claims 1-164.
168. A method comprising: administering to the subject a pan-*coronavirus* recombinant vaccine composition according to any of claims 1-164, wherein the composition prevents virus replication in the lungs, the brain, and other compartments where the virus replicates.
169. A method comprising: administering to the subject a pan-*coronavirus* recombinant vaccine composition according to any of claims 1-164, wherein the composition prevents cytokine storm in the lungs, the brain, and other compartments where the virus replicates.
170. A method comprising: administering to the subject a pan-*coronavirus* recombinant vaccine composition according to any of claims 1-164, wherein the composition prevents inflammation or inflammatory response in the lungs, the brain, and other compartments where the virus replicates.

171. A method comprising: administering to the subject a pan-coronavirus recombinant vaccine composition according to any of claims 1-164, wherein the composition improves homing and retention of T cells in the lungs, the brain, and other compartments where the virus replicates.
172. A method of preventing a *coronavirus* disease in a subject; the method comprising: administering to the subject a pan-coronavirus recombinant vaccine composition according to any of claims 1-164, wherein the composition induces memory B and T cells.
173. The method of claim 172, wherein the composition induces resident memory T cells (Trm).
174. A method of prolonging an immune response induced by a pan-*coronavirus* vaccine and increasing T-cell migration to the lungs, the method comprising: co-expressing a T-cell attracting chemokine, a composition that promotes T cell proliferation, and a pan-*coronavirus* recombinant vaccine composition according to any of claims 1-164.
175. A method of prolonging the retention of memory T-cell into the lungs induced by a pan *coronavirus* vaccine and increasing virus-specific tissue resident memory T-cells (T_{RM} cells), the method comprising: co-expressing a T-cell attracting chemokine, a composition that promotes T cell proliferation, and a pan-*coronavirus* vaccine according to any of claims 1-164.
176. A method comprising: administering to the subject a pan-coronavirus recombinant vaccine composition according to any of claims 1-164, wherein the composition prevents the development of mutation and variants of a *coronavirus*.
177. The method of any of claims 165-176, wherein the vaccine is administered through an intravenous route (i.v.), an intranasal route (i.n.), or a sublingual route (s.l.) route.
178. The method of any of claims 165-177, wherein the recombinant vaccine composition induces efficient and powerful protection against the *coronavirus* disease or infection.
179. The method of any of claims 165-178, wherein the recombinant vaccine composition induces production of antibodies (Abs), CD4+ T helper (Th1) cells, and CD8+ cytotoxic T-cells (CTL).
180. The method of any of claims 165-179, wherein one or more of the T-cell attracting chemokine, the composition that promotes T cell proliferation, and the pan-*coronavirus* vaccine is operatively linked to a promoter.
181. The method of claim 180, wherein the promoter is a lung specific promoter.
182. The method of claim 181, wherein the lung specific promoter is SP-B or CD144.
183. The method of claim 180, wherein the promoter is a generic promoter.
184. The method of claim 183, wherein the promoter is a human cytomegalovirus immediate early enhancer/promoter (CMV).
185. The method of any of claims 165-185, wherein the composition that promotes T cell proliferation is IL-7, IL-2, or IL-15.
186. The method of any of claims 165-186, wherein the composition that promotes T cell proliferation helps to promote long term immunity.
187. The method of any of claims 165-187, wherein the T cell attracting chemokine is CCL5, CXCL9, CXCL10, CXCL11, or a combination thereof.
188. The method of any of claims 165-188, wherein the T-cell attracting chemokine helps pull T-cells from circulation into the lungs.

189. A method comprising:
- administering a first dose of a pan-coronavirus vaccine composition according to any of claims 1-164 using a first delivery system; and
 - administering a second dose using a second delivery system; wherein the first and second delivery system are different.
190. The method of claim 189, wherein the first delivery system may comprise a RNA, a modified mRNA, or a peptide delivery system.
191. The method of claim 189-190, wherein the second delivery system may comprise a RNA, a modified mRNA, or a peptide delivery system.
192. The method of claim 189, wherein the peptide delivery system is an adenovirus or an adeno-associated virus.
193. The method of claim 190, wherein the adenovirus delivery system is Ad26, Ad5, Ad35, or a combination thereof.
194. The method of claim 190, wherein the adeno-associated delivery system is AAV8 or AAV9.
195. The method of any of claims 189-194 wherein the second vaccine dose is administered 14 days after the first vaccine dose.
196. A method comprising:
- administering a pan-coronavirus recombinant vaccine composition according to any of claims 1-164; and
 - administering at least one T-cell attracting chemokine after administering the pan-coronavirus recombinant vaccine composition.
197. The method of claim 196, wherein the recombinant vaccine composition is administered via a RNA, a modified mRNA, or a peptide delivery system.
198. The method of claim 196-197, wherein the T-cell attracting chemokine is administered via a RNA, a modified mRNA, or a peptide delivery system.
199. The method of any of claims 196-198 wherein the peptide delivery system is an adenovirus or an adeno-associated virus.
200. The method of claim 199, wherein the adenovirus delivery system is Ad26, Ad5, Ad35, or a combination thereof.
201. The method of claim 199, wherein the adeno-associated delivery system is AAV8 or AAV9.
202. The method of any of claims 196-201, wherein the T-cell attracting chemokine is administered 14 days after administering days after the recombinant vaccine composition.
203. The method of any of claims 196-202, wherein the T-cell attracting chemokine is CCL5, CXCL9, CXCL10, CXCL11, or a combination thereof.
204. A method comprising:
- administering a pan-coronavirus recombinant vaccine composition according to any of claims 1-164;
 - administering at least one T-cell attracting chemokine after administering the pan-coronavirus recombinant vaccine composition; and
 - administering at least one cytokine after administering the T-cell attracting chemokine.

205. The method of claim 204, wherein the recombinant vaccine composition is administered via a RNA, a modified mRNA, or a peptide delivery system.
206. The method of any of claims 204-205, wherein the T-cell attracting chemokine is administered via a RNA, a modified mRNA, or a peptide delivery system.
207. The method of any of claims 204-206, wherein the cytokine is administered via a RNA, a modified mRNA, or a peptide delivery system.
208. The method of any of claims 204-207, wherein the peptide delivery system is an adenovirus or an adeno-associated virus.
209. The method of claim 208, wherein the adenovirus delivery system is Ad26, Ad5, Ad35, or a combination thereof.
210. The method of claim 208, wherein the adeno-associated delivery system is AAV8 or AAV9.
211. The method of any of claims 204-210, wherein the T-cell attracting chemokine is administered 14 days after administering the recombinant vaccine composition.
212. The method of any of claims 204-211, wherein the T-cell attracting chemokine is CCL5, CXCL9, CXCL10, CXCL11, or a combination thereof.
213. The method of any of claims 204-212, wherein the cytokine is administered 10 days after administering the T-cell attracting chemokine.
214. The method of claim 16, wherein the cytokine is IL-7, IL-15, or a combination thereof.
215. A method comprising:
- a) administering a pan-coronavirus recombinant vaccine composition according to any of claims 1-164;
 - b) administering one or more T-cell attracting chemokine after administering the pan-coronavirus recombinant vaccine composition; and
 - c) administering one or more mucosal chemokines.
216. The method of claim 215, wherein the recombinant vaccine composition is administered using modified RNA, adeno-associated virus, or an adenovirus.
217. The method of any of claims 215-216, wherein the T-cell attracting chemokine is administered via a RNA, a modified mRNA, or a peptide delivery system.
218. The method of any of claims 215-216, wherein the mucosal chemokine is administered via a RNA, a modified mRNA, or a peptide delivery system.
219. The method of claim 218, wherein the adeno-associated virus is AAV8 or AAV9.
220. The method of claim 218, wherein the adenovirus is Ad26, Ad5, Ad35, or a combination thereof.
221. The method of any of claims 215-220, wherein the T-cell attracting chemokine is administered 14 days after administering the recombinant vaccine composition.
222. The method of any of claims 215-221, wherein the T-cell attracting chemokine is CCL5, CXCL9, CXCL10, CXCL11, or a combination thereof.
223. The method of any of claims 215-222, wherein the mucosal chemokine is administered 10 days after administering the T-cell attracting chemokine.
224. The method of any of claims 215-223, wherein the mucosal chemokine is CCL25, CCL28,

CXCL14, or CXCL17, or a combination thereof.

225. The method of any of claims 215-224, wherein the recombinant vaccine composition is administered via two doses.
226. The method of any of claims 215-225, wherein the doses are about 3 weeks apart.
227. A pan-coronavirus recombinant vaccine composition, the composition comprising one or more large sequences, each of the one or more large sequences comprises at least one of:
- a) whole spike protein or a portion thereof;
 - b) one or more conserved *coronavirus* CD4⁺ T cell target epitope; and
 - c) one or more conserved *coronavirus* CD8⁺ T cell target epitope;
- wherein at least one epitope is derived from a non-spike protein.
228. The composition of claim 227, wherein the one or more conserved epitopes are highly conserved among human and animal *coronaviruses*.
229. The composition of any of claims 227-228, wherein the one or more conserved epitopes are derived from at least one of SARS-CoV-2 protein.
230. The composition of any of claims 227-229, wherein the composition comprises 2-20 CD8⁺ T cell target epitopes.
231. The composition of any of claims 227-230, wherein the composition comprises 2-20 CD4⁺ T cell target epitopes.
232. The composition of any of claims 227-231, wherein the one or more conserved *coronavirus* CD4⁺ T cell target epitopes selected from SEQ ID NO: 58-105 (ORF1a₁₃₅₀₋₁₃₆₅, ORF1ab₅₀₁₉₋₅₀₃₃, ORF6₁₂₋₂₆, ORF1ab₆₀₈₈₋₆₁₀₂, ORF1ab₆₄₂₀₋₆₄₃₄, ORF1a₁₈₀₁₋₁₈₁₅, S₁₋₁₃, E₂₆₋₄₀, E₂₀₋₃₄, M₁₇₆₋₁₉₀, N₃₈₈₋₄₀₃, ORF7a₃₋₁₇, ORF7a₁₋₁₅, ORF7b₈₋₂₂, ORF7a₉₈₋₁₁₂, and ORF8₁₋₁₅).
233. The composition of any of claims 227-232, wherein the one or more conserved *coronavirus* CD8⁺ T cell target epitopes selected from SEQ ID NO: 106-138 (S₂₈₇₋₃₁₇, S₅₂₄₋₅₉₈, S₆₀₁₋₆₄₀, S₈₀₂₋₈₁₉, S₈₈₈₋₉₀₉, S₃₆₉₋₃₉₃, S₄₄₀₋₅₀₁, S₁₁₃₃₋₁₁₇₂, S₃₂₉₋₃₆₃, and S₁₃₋₃₇).
234. A pan-coronavirus recombinant vaccine composition, the composition comprising one or more large sequences, each of the one or more large sequences comprises at least one of:
- a) one or more conserved *coronavirus* B-cell target epitope;
 - b) one or more conserved *coronavirus* CD4⁺ T cell target epitope; and/or
 - c) one or more conserved *coronavirus* CD8⁺ T cell target epitope;
- wherein at least one epitope is derived from a non-spike protein.
235. The composition of claim 234, wherein the one or more conserved epitopes are derived from at least one of SARS-CoV-2 protein.
236. The composition of any of claims 234-235, wherein the composition comprises 2-20 CD8⁺ T cell target epitopes.
237. The composition of any of claims 234-236, wherein the composition comprises 2-20 CD4⁺ T cell target epitopes.
238. The composition of any of claims 234-237, wherein the one or more conserved *coronavirus* CD4⁺ T cell target epitopes selected from SEQ ID NO: 58-105 (ORF1a₁₃₅₀₋₁₃₆₅, ORF1ab₅₀₁₉₋₅₀₃₃, ORF6₁₂₋₂₆, ORF1ab₆₀₈₈₋₆₁₀₂, ORF1ab₆₄₂₀₋₆₄₃₄, ORF1a₁₈₀₁₋₁₈₁₅, S₁₋₁₃, E₂₆₋₄₀, E₂₀₋₃₄, M₁₇₆₋₁₉₀, N₃₈₈₋₄₀₃,

- ORF7a₃₋₁₇, ORF7a₁₋₁₅, ORF7b₈₋₂₂, ORF7a₉₈₋₁₁₂, and ORF8₁₋₁₅).
239. The composition of any of claims 234-238, wherein the one or more conserved *coronavirus* CD8⁺ T cell target epitopes selected from SEQ ID NO: 106-138 (S₂₈₇₋₃₁₇, S₅₂₄₋₅₉₈, S₆₀₁₋₆₄₀, S₈₀₂₋₈₁₉, S₈₈₈₋₉₀₉, S₃₆₉₋₃₉₃, S₄₄₀₋₅₀₁, S₁₁₃₃₋₁₁₇₂, S₃₂₉₋₃₆₃, and S₁₃₋₃₇).
240. The composition of any of claims 234-239, wherein the one or more conserved *coronavirus* B-cell target epitopes selected from SEQ ID NO: 2-57 (S₂₋₁₀, S₁₂₂₀₋₁₂₂₈, S₁₀₀₀₋₁₀₀₈, S₉₅₈₋₉₆₆, E₂₀₋₂₈, ORF1ab₁₆₇₅₋₁₆₈₃, ORF1ab₂₃₆₃₋₂₃₇₁, ORF1ab₃₀₁₃₋₃₀₂₁, ORF1ab₃₁₈₃₋₃₁₉₁, ORF1ab₅₄₇₀₋₅₄₇₈, ORF1ab₆₇₄₉₋₆₇₅₇, ORF7b₂₆₋₃₄, ORF8a₇₃₋₈₁, ORF10₃₋₁₁, and ORF10₅₋₁₃);
241. A pan-coronavirus recombinant vaccine composition, the composition comprising an antigen delivery system encoding one or more large sequences, the large sequences comprise at least one of:
- a) one or more conserved *coronavirus* B-cell target epitopes;
 - b) one or more conserved *coronavirus* CD4⁺ T cell target epitopes; and/or
 - c) one or more conserved *coronavirus* CD8⁺ T cell target epitopes;
- wherein at least one epitope is derived from a non-spike protein.
242. The composition of claim 241, wherein the antigen delivery system is an adenovirus-based antigen delivery system.
243. The composition of claim 242, wherein the adenovirus-based antigen delivery system is Ad26, Ad5, Ad35, or a combination thereof.
244. The composition of any of claims 241-243, wherein the antigen delivery system further encodes a T cell attracting chemokine.
245. The composition of any of claims 241-244, wherein the antigen delivery system further encodes a composition that promotes T cell proliferation.
246. The composition of any of claims 241-245, wherein the antigen delivery system further encodes a molecular adjuvant.
247. The composition of any of claims 241-246, wherein the epitopes are operatively linked to a lung-specific promoter.
248. The composition of any of claims 241-247, wherein the one or more conserved *coronavirus* B-cell target epitopes selected from SEQ ID NO: 2-57 (S₂₋₁₀, S₁₂₂₀₋₁₂₂₈, S₁₀₀₀₋₁₀₀₈, S₉₅₈₋₉₆₆, E₂₀₋₂₈, ORF1ab₁₆₇₅₋₁₆₈₃, ORF1ab₂₃₆₃₋₂₃₇₁, ORF1ab₃₀₁₃₋₃₀₂₁, ORF1ab₃₁₈₃₋₃₁₉₁, ORF1ab₅₄₇₀₋₅₄₇₈, ORF1ab₆₇₄₉₋₆₇₅₇, ORF7b₂₆₋₃₄, ORF8a₇₃₋₈₁, ORF10₃₋₁₁, and ORF10₅₋₁₃).
249. The composition of any of claims 241-248, wherein the one or more conserved *coronavirus* CD4⁺ T cell target epitopes selected from SEQ ID NO: 58-105 (ORF1a₁₃₅₀₋₁₃₆₅, ORF1ab₅₀₁₉₋₅₀₃₃, ORF6₁₂₋₂₆, ORF1ab₆₀₈₈₋₆₁₀₂, ORF1ab₆₄₂₀₋₆₄₃₄, ORF1a₁₈₀₁₋₁₈₁₅, S₁₋₁₃, E₂₆₋₄₀, E₂₀₋₃₄, M₁₇₆₋₁₉₀, N₃₈₈₋₄₀₃, ORF7a₃₋₁₇, ORF7a₁₋₁₅, ORF7b₈₋₂₂, ORF7a₉₈₋₁₁₂, and ORF8₁₋₁₅).
250. The composition of any of claims 241-249, wherein the one or more conserved *coronavirus* CD8⁺ T cell target epitopes selected from SEQ ID NO: 106-138 (S₂₈₇₋₃₁₇, S₅₂₄₋₅₉₈, S₆₀₁₋₆₄₀, S₈₀₂₋₈₁₉, S₈₈₈₋₉₀₉, S₃₆₉₋₃₉₃, S₄₄₀₋₅₀₁, S₁₁₃₃₋₁₁₇₂, S₃₂₉₋₃₆₃, and S₁₃₋₃₇).
251. The composition of any of claims 241-250, comprising a spike protein or a partial spike protein.

252. The composition of claim 251, wherein the partial spike protein comprises a trimerized SARS-CoV-2 receptor-binding domain (RBD).
253. The composition of any of claims 251-252, wherein the whole spike protein or partial spike protein has an intact S1-S2 cleavage site.
254. The composition of any of claims 251-253, wherein the spike protein is stabilized with proline substitutions at amino acid positions 986 and 987.
255. A pan-coronavirus recombinant recombinant vaccine composition comprising one of SEQ ID NO: 139-147.