

T CELL BROAD BETACORONAVIRUS VACCINE

FIELD OF THE INVENTION

The present invention relates to polypeptides, polynucleotides, compositions, microorganisms, vectors and vaccine compositions optimised for the prophylactic or therapeutic treatment of an infection caused by Betacoronaviruses, including but not limited to: Embecovirus, Hibecovirus, Merbecovirus, Nobecovirus, Sarbecovirus, MERS-CoV, SARS-CoV-1 and SARS-CoV-2.

BACKGROUND

The outbreak of coronavirus disease 2019 (COVID-19) and its rapid worldwide transmission resulted in its declaration as a pandemic and global health emergency by the World Health Organisation (WHO). COVID-19 is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), a positive-sense RNA coronavirus that has an envelope encapsulating its large RNA genome and is further characterised by an exposed spike glycoprotein (S-protein), projecting from its viral surface (Gorbalenya *et al.* 2020, Nat Microbiol 5(4): 536-544). The widespread proliferation of SARS-CoV-2 follows previous deadly outbreaks of SARS-CoV-1 and MERS-CoV, of which all three are subgenera of the genus Betacoronavirus.

Whilst the majority of COVID-19 cases result only in mild symptoms including fever, cough, or shortness of breath, a significant number of cases progress to viral pneumonia and multi-organ failure (Hui *et al.* 2020, Int J Infect Dis 91: 264-66). The rapid rise in the number of infections and deaths around the globe highlights the urgent need for better therapeutic and prophylactic interventions to combat the disease and an effective vaccine has been hailed by many as a crucial cornerstone in our potential fight against the SARS-CoV-2 virus. Vaccines for SARS-CoV-1 and MERS-CoV were also in development before the respective outbreaks of the viruses came to an end, removing the need for vaccines (Tseng *et al.*, 2012, PLoS ONE, 7(4): e35421; Modjarrad *et al.*, 2019, The Lancet

Infectious Diseases, 19(9), 1013-1022). However, it is clear from the widespread outbreaks of MERS-CoV, SARS-CoV-1 and SARS-CoV-2 that viruses of the Betacoronavirus genus present a critical threat to humanity which requires both prophylactic and therapeutic treatment.

- 5 Vaccination has been established as an effective form of epidemiological control, and vaccines have had significant success in aiding the decline of infections and mortalities associated with viral infections such as smallpox and polio. Other infections, however, have proven harder to vaccinate against. A further challenge exists in the creation of vaccines that provide protection against multiple
- 10 pathogenic species. Due to the genetic differences that exist between even closely related pathogens, the creation of a single vaccine that may provide protection against many species is especially difficult. Furthermore, quickly evolving pathogens such as viruses may evolve to produce a variety of strains, as seen during the COVID-19 pandemic, wherein the original vaccine no longer provides
- 15 the required protection. As such, searching for regions of pathogens' genomes that are highly conserved either between species or strains may constitute a strategy for selecting broadly protective vaccines.

- As one arm of the adaptive immune system that is specialised to resolve infections and prevent reinfection from pathogens, cellular immunity often works in tandem
- 20 with humoral – antibody-based – immunity upon natural exposure to a foreign body. A cellular immune response involves the interaction of T cells, each providing a variety of immune-related functions to aid in the reduction or elimination of pathogen-infected host cells (Amanna & Slifka 2011, Virology 411(2): 206-215). T-cell immunity has been shown to play a role in limiting the
- 25 acquisition of symptomatic SARS-COV-2 infection through clearance of infected cells by the cellular adaptive response, which limits replication and spreading of the virus to further cells and tissues (Kent et al. 2022, Nature Reviews Immunology 22, 387-397). Furthermore, the generation of memory T cells as part of the cellular immune response results in the ability to mount a faster and stronger immune
- 30 response upon re-exposure to a previously encountered pathogen (Restifo &

Tattinoni 2013, Current Opinion in Immunology 25(5): 556-63). A T-cell response might be targeted to any protein of a viral pathogen, while a humoral response is usually directed to surface proteins. This allows for the identification of a relatively large number of targets for each pathogen. As SARS-CoV-2 vaccine development
5 has been focused on activating a neutralising antibody-based humoral immune response, most commonly through the generation of S protein-based subunit vaccines (Amanat & Krammer 2020, Cell Press Immunity 52: 583 – 589), however such subunit vaccines are unlikely to generate robust cellular immune responses in a broad population (Testa & Philip 2012, Future Virol 7(11): 1077-1088).

10 However, when designing vaccines engineered to induce a broad T cell response, there exists a further challenge of human leukocyte antigen (HLA) restriction within an individual and a broader population. The HLA complex is a set of genes encoding the major histocompatibility complex (MHC) proteins in humans, responsible for the regulation of an individual's immune system, as well as the
15 ability to specifically present at the surface of infected cells, and elicit an immune response against, epitopes generated during a natural infection, or delivered to said individual in the form of a vaccine (Marsh *et al.* 2010 Tissue Antigens 75(4): 291-455).

The high polymorphism of HLA alleles and subsequent immune system variability
20 between individuals results in a diverse spectrum of "HLA types" across the population. As an added complication to peptide-based vaccine development, such HLA types can have a significant impact on the efficacy of a potentially prophylactic viral vaccine composition between different individuals. As such, generation of an epitope-based vaccine composition that is compatible with a
25 particular subset of HLA types may prove ineffective with a significant proportion of the global population comprising individuals of different HLA types. In light of this, the generation of T-cell and B-cell epitope vaccines, that target a limited number of HLA types, may only prove advantageous for a narrow, select population.

The current lack of an approved vaccine composition which is efficacious across a wide range of HLA populations, creates significant danger for at-risk populations, including health care workers and patients in acute danger of nosocomial or community-transmitted infections.

- 5 Thus, there exists an urgent need for a safe and effective vaccine for use in the therapeutic or prophylactic treatment of Betacoronavirus subgenera, optimised to incorporate epitopes covering a diverse spectrum of HLA types, with the potential to stimulate a broad adaptive immune response against Betacoronavirus subgenera across the global human population.

10 SUMMARY OF INVENTION

- This invention is based on the surprising discovery that, by using a platform to identify predicted cross-reactive Betacoronavirus epitopes that bind HLA molecules across a broad spectrum of HLA types, potentially broad spectrum vaccines with predicted high efficacy can be formulated that comprise one or more
15 of said epitopes. Such a vaccine thus has the potential to stimulate a broad adaptive immune response to Betacoronaviruses that is both cellular and potentially humoral in nature, for the therapeutic or prophylactic treatment of infections with Betacoronaviruses in humans across the global population.

- In a first aspect of the invention, there is provided a vaccine composition
20 comprising a polypeptide, wherein the polypeptide comprises one or more epitope sequences, wherein the one or more epitope sequences have the amino acid sequences of any one or more of SEQ ID Nos 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 24, 26, 28, 30, 31, 33, 35, 37, 41, 45, 46, 48, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 91, 93, 94, 96, 98, 100, 102, 104, 106,
25 108, 110, 113, 114, 116, 118, 120, 124, 126, 128, 130, 132, 134, 137, 143, 161, 165, 169, 176, 182, 205, 209, 219, 229, 231, 233 and/or 241, or a variant thereof having at least 70% sequence identity thereto, and wherein the polypeptide sequence is no more than 1400 amino acids in length.

It is envisaged that the polypeptide may, for example, comprise at least 2, 3, 4, 5, 6, 7, 8, 9, 10 or more epitope sequences having the amino acid sequences of any of the sequences of claim 1. It is also envisaged that the polypeptide may comprise a concatemer of epitope sequences each having the amino acid sequences of claim 1. It is envisaged that the polypeptide may comprise any of the following groups of epitope sequences: a. SEQ ID No 1 to 21, or b. SEQ ID No 22 to 93, or c. SEQ ID No 94 to 113, or d. SEQ ID No 114 to 134, or e. SEQ ID No 135 to 152, or f. SEQ ID No 153 to 190, or g. SEQ ID No 191 to 208, or h. SEQ ID No 209 to 245. It is also envisaged that the polypeptide may comprise any of the following groups of epitope sequences arranged in the following order (in other words, in numerical order arranged in series) on the polypeptide: a. SEQ ID No 1 to 21, or b. SEQ ID No 22 to 93, or c. SEQ ID No 94 to 113, or d. SEQ ID No 114 to 134, or e. SEQ ID No 135 to 152, or f. SEQ ID No 153 to 190, or g. SEQ ID No 191 to 208, or h. SEQ ID No 209 to 245. By "concatemer" we intend that the polypeptide may comprise multiple copies of an indicated epitope sequence linked in series. The number of copies will be dependent on the requirements of the polypeptide but it may be for example, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more copies of the epitope in question. The polypeptide may further comprise one or more spacer regions. It is envisaged that the spacer regions may be positioned between the epitope sequences. Although, as the skilled person will appreciate, depending on the adjacent epitope sequences in question, a spacer may not be required. In the instance of the polypeptide being a concatemer, the polypeptide may optionally include spacer regions, as would be determined by a person of skill in the art. One consideration here is whether undesirable epitopes may be created at the junction of two epitopes when a spacer is not included. Thus, in some embodiments, some epitopes may be separated by a spacer, while others are not. Equally, all epitopes may be separated by spacer regions. Thus, when the polypeptide comprises more than one of the epitope sequences of the invention or a variant thereof, the one or more spacer regions may be positioned between one or more or each epitope sequence. In an embodiment of the invention, each of the one or more epitope sequences of the polypeptide have at least 70%, 80%,

90%, 95%, 96%, 97%, 98% or 99% sequence identity to any of the sequences of the invention. In a further embodiment of the invention, the polypeptide is no more than 1400, 1300, 1200, 1100, 1000, 900, 800, 700, 600, 500, 400, 300, 200, 150, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10 or 9 amino acids in length.

- 5 In a second aspect of the invention, there is provided a vaccine composition according to the first aspect of the invention, for therapeutic or prophylactic use.

In a third aspect of the invention, there is provided a vaccine composition according to the first aspect of the invention, for use in the treatment or prophylaxis of an infection caused by a Betacoronavirus.

- 10 In a fourth aspect of the invention, there is provided a use of a vaccine composition according to the first aspect of the invention, in the manufacture of a medicament for the treatment of a Betacoronavirus infection.

- In a fifth aspect of the invention, there is provided a method of treating a Betacoronavirus infection, comprising administering to a subject an effective
15 amount of the vaccine composition according to the first aspect of the invention.

- In a sixth aspect of the invention, there is provided a vector encoding the polypeptide as defined in the first aspect of the invention, preferably wherein the vector may further comprise regulatory elements capable of driving transcription and/or translation of the polynucleotide in a host cell. The vector may be a viral
20 vector or any other suitable vector.

In a seventh aspect of the invention, there is provided a microorganism comprising a polypeptide as defined in the first aspect of the invention, or a polynucleotide encoding the polypeptide as defined in the first aspect of the invention. The microorganism may be a bacterial microorganism.

- 25 In an eighth aspect of the invention, there is provided a polynucleotide as defined in the first aspect of the invention, or a vector according to the fifth aspect of the invention, formulated in a nanoparticle or lipid formulation.

In a ninth aspect of the invention, there is provided a composition comprising:

i) one or more epitope sequences comprising any one or more of SEQ ID Nos 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 24, 26, 28, 30, 31, 33, 35, 37, 41, 45, 46, 48, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 91, 93, 94, 96, 98, 100, 102, 104, 106, 108, 110, 113, 114, 116, 118, 120, 124, 126, 128, 130, 132, 134, 137, 143, 161, 165, 169, 176, 182, 205, 209, 219, 229, 231, 233 and/or 241, or a variant thereof having at least 70% sequence identity, each of said epitope sequences or variants thereof being no more than 1400 amino acids in length, or

ii) one or more polynucleotides encoding one or more epitope sequences comprising any one or more of SEQ ID Nos 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 24, 26, 28, 30, 31, 33, 35, 37, 41, 45, 46, 48, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 91, 93, 94, 96, 98, 100, 102, 104, 106, 108, 110, 113, 114, 116, 118, 120, 124, 126, 128, 130, 132, 134, 137, 143, 161, 165, 169, 176, 182, 205, 209, 219, 229, 231, 233 and/or 241, or a variant thereof having at least 70% sequence identity, each of said epitope sequences or variants thereof being no more than 1400 amino acids in length.

In a tenth aspect of the invention, there is provided a vector, microorganism or composition according to the above aspects of the invention for therapeutic or prophylactic use.

In an eleventh aspect of the invention, there is provided a vector, microorganism or composition according to the above aspects of the invention, for use in the treatment or prophylaxis of an infection caused by a Betacoronavirus.

In a twelfth aspect of the invention, there is provided a vector, microorganism or composition for use according to the eleventh aspect of the invention, wherein the infection is caused by a member of the Embecovirus, Hibecovirus, Merbecovirus, Nobecovirus or Sarbecovirus subgenera, or by MERS-CoV, SARS-CoV or SARS-CoV-2.

BRIEF DESCRIPTION OF DRAWINGS

- Figure 1** shows the hotspot detection workflow. Using the Antigen Presentation (AP) score map of a given protein (1), where each row represents one of the 150 considered Class I HLA alleles and each column represents an amino acid within the protein sequence, the sum of AP scores across HLAs is calculated for each amino acid position (2), as well as the mean of these summed up AP scores (dashed line). The mean of the AP score sums is used as a baseline and a LOESS smoothing curve is applied on the AP score sums at amino acid level (3). Next, the peak summits along the smoothing curve are identified using a rolling maximum function (circles) and the peak regions identified as windows of consecutive amino acids consistently above the baseline (rectangles). Finally, the hotspot sequences of the peak regions are extracted, as well as their start and end coordinates within the protein sequence (4).
- Figure 2** shows the conserved region identification flow. The Clustal Omega output of the re-aligned extended hotspots (A) is used to extract the sequence of consensus symbols (rectangle in A). Using run length encoding (RLE), conserved hotspot regions are identified as consecutive amino acids denoted by an "*" or "." within the Clustal Omega consensus symbol sequence (rectangles in B).
- Figure 3** shows the candidate minimal epitope immunogenicity analysis. The immunogenicity corresponds to the Antigen Presentation (AP) score of the epitope. Each dot corresponds to one candidate minimal epitope. A handful of epitopes are present in all five Betacoronavirus subgenera. The Figure shows the level of conservation of a candidate minimal epitope across the tested Betacoronavirus subgenera. The more conserved a minimal candidate epitope is, the higher immunogenicity it will have across more subgenera of Betacoronavirus.
- Figure 4** shows the population coverage by HLA-A frequency for a particular optimized Betacoronavirus vaccine design. There is a tradeoff between population

coverage and Betacoronavirus subgenera coverage. Generally, as more of the human population would respond to the broadly protective vaccine, it protects against fewer Betacoronavirus subgenera.

Figure 5 shows the amino acid sequences of SEQ ID Nos 1-245 grouped into eight polypeptides that are predicted to be cross-reactive across the five Betacoronavirus subgenera and across multiple HLAs.

Figure 6 shows the procedure for testing CD8+ and CD4+ T cell responses to minimal epitope Betacoronavirus peptide pools or control peptide pools after generation of T cell lines (*ex vivo* vaccination).

Figure 7 shows analysis of T cell response frequency after challenging T cell lines with minimal epitope Betacoronavirus peptide pools or control peptide pools. The figure graphically presents the polyfunctionality distributions of responses of CD8+ (top) and CD4+ (bottom) T cells. Response was measured by the magnitude of the increase above background expression of the activation markers CD40L, CD137, IFNG, and TNF measured by flow cytometry. Expression marker permutations are shown in grey, corresponding relative percentages of the T-cell populations defined by those markers are shown for each donor.

Figure 8 shows the numerical polyfunctionality response of CD8+ and CD4+ T cells from T cell lines as shown graphically in Figure 9. (A) Total polyfunctional response (increase above background) of expanded CD8+ and CD4+ T cells as measured in Figure 7. (B-C) Fraction of CD8+(B) and CD4+ (C) T-cell subpopulations defined by 4 activation markers measured by flow (CD137, CD40L, CD137, IFNG, and TNF). The increase in quadruple, triple, and double positive populations indicates polyfunctional response. Distribution of markers is shown by Arcs in pie charts and subpopulation legends. Donor ID is reported below each pie chart.

Figure 9 shows a graphical presentation of the polyfunctionality distributions of responses of CD8+ and CD4+ T cells in a T cell line stimulated with peptides

derived from construct t_cell_d5-mm20-620-wf, See Figure 5. (A) Total polyfunctional response (increase above background) of expanded CD8+ and CD4+ T cells as measured in Figure 8. (B) Fraction of CD8+ and CD4+ T-cell subpopulations defined by 4 activation markers measured by flow (CD40L, CD137, IFNG, and TNF). The increase in quadruple, triple, and double positive populations indicates polyfunctional response. Distribution of markers is shown by Arcs in pie charts and subpopulation legends. Donor ID is reported below each pie chart.

Figure 10 shows CD4+ and CD8+ T cell Reactivity scores (Federico et al, Frontiers Immun. Nov 2023) directly ex vivo in response to single immunoreactive Betacoronavirus peptides used as listed in Table 5. (A-B) Reactivity scores (Federico et al, Frontiers Immun. Nov 2023) of vaccine peptides in 7 healthy donors who received ≥ 2 doses of (mRNA) vaccines (Moderna/mRNA-1273 or Pfizer/BioNTech BNT162b2) and had undergone COVID-19. Single peptide IDs shown in the graphs for each patients are the following: HD110 (sp4-10-13-25-38-39-41-42-48-50-74-81-102), HD112 (sp5-15-24-30-42-43-49-52-56-59-62-81-93-102-106), HD114 (sp15-24-42-62-81-93-102-106), HD117 (sp10-13-38-39-41-42-48-50-81-102), HD118 (sp5-10-13-24-30-38-39-41-43-48-49-50-52-56-59-62), HD131 (sp4-5-25-30-43-49-52-56-59-74), HD138 (sp4-5-74) for both the CD8+ (A) and CD4+ (B) T-cell subsets following (see Table 5 for details). Low scoring peptides in any T-cell subset are indicated.

DETAILED DESCRIPTION

This invention is predicated on the development of a platform that is able to identify Betacoronavirus epitopes that have a high probability of effectively stimulating a broad adaptive cellular immune response to multiple Betacoronavirus variants when used in a vaccine composition. Thus, the incorporation of such epitopes into peptides, polynucleotides, compositions, microorganisms, vectors or a vaccine composition may allow for the therapeutic or prophylactic treatment of infection with Betacoronavirus. Unlike prior Betacoronavirus vaccination

approaches, the polypeptide, polynucleotide, vector, microorganism, composition or vaccine composition of the present invention is designed to stimulate a broad adaptive immune response through the specific activation of CD8+ and CD4+ T cells. It is expected that this will lead to the generation of a more substantial and sustained level of immunity. Furthermore, a surprisingly robust statistical model allows for the identification of those predicted Betacoronavirus epitopes that are capable of triggering immunogenicity across a wide variety of human leukocyte antigen (HLA) types, hence the vaccine composition may have the potential to elicit protection against Betacoronavirus across the global human population.

5 Through a combined analysis of predicted antigen presentation (AP) scores and HLA types, amino acid sequences that are highly conserved across multiple Betacoronavirus subgenera, as well as comprising likely immunogenic epitopes across a variety of HLA types representative of the diversity found in a human population, were identified. In this way, the platform upon which the present invention is predicated identifies desirable Betacoronavirus epitopes capable of eliciting an immune response against a variety of Betacoronavirus subgenera, across a wide variety of HLA types. The conserved (epitope) sequences are also referred to herein as "hotspot" regions or sequences.

10

15

AP scores were initially used to identify regions of the genomes of Betacoronaviruses that contained hotspot regions, which in turn comprise likely immunogenic epitopes. These regions were allocated AP scores between 0 and 1, wherein a higher AP score indicated a higher likelihood of the presence of immunogenic epitopes (Figure 1). Once the regions were identified, the AP scores had a further use in identifying two types of desirable vaccine candidates. The first concerned AP scores or sums of AP scores of hotspot regions that contained a single or low number of epitopes that were immunogenic in a single or low number of HLA types. These AP scores were close to 1 (Table 3). The second concerned the sum of AP scores for multiple epitopes that appear within a given hotspot region. By adding the values of the individual AP scores (between 0 and 1) together, hotspot regions that are enriched with immunogenic epitopes were

20

25

30

identified. As the AP scores of the epitopes appearing within the hotspot region were added together, the sum of the AP scores is comparatively high to the first type of desirable candidate – appearing in the region of 10 to 80. These sums of AP scores are presented in Table 3. Through this methodology, the inventors were able to first identify regions of Betacoronavirus genomes that contained more likely immunogenic epitopes than the rest of the genomes, and then analyse hotspot regions and the epitopes that comprise them for desirable vaccine candidates.

In a first aspect of the invention, there is provided a vaccine composition comprising a polypeptide, wherein the polypeptide comprises one or more epitope sequences, wherein the one or more epitope sequences have the amino acid sequences of any one of the sequences in Table 1 below, or a variant thereof having at least 70% sequence identity thereto, and wherein the polypeptide sequence is no more than 1400 amino acids in length.

The amino acid sequences of the invention are as follows in Table 1:

Table 1: Amino acid hotspot sequences identified to be broadly immunogenic across multiple Betacoronavirus subgenera and across multiple HLAs.

SEQ ID No	Amino acid sequence
1	FVDGVPFVVSIGYHYKELGIVMNMDVDTHRYRLSLKDLLLYAADP
3	NRATLQAIASEFSSLPSYAAAYATAQEAY
5	QRVFNYPYVFTLLFQLCTFTKSTNSRIRASLPTTIKN
7	AEHFIVTVSLAGSYKDWSSGQRTELGVFLKRGDKIVYHTVGNVIEFH MEGEVLPLEKLTLLALREVKTIVFTTVDNINLHTQVIDMSMTYGGQLG
9	LYRLVVDAYQAPPTI
11	GFSKFVTVQGPPGTGKSHFAIGLAIYYPTARVY
13	LHVTYVPSQERNFTTAPAICHEGKAYFPREGVFVF
15	IMHVSFLVMFSGVVPLWTVLYVVVVLGRHVFWFTSLCTKRTVHVGDV FHSFQDAALQTFMLDKEVFLRLKREISPETYARYLALFNKYKYSGPMDT A
17	TSFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVNLHSSRLSFKELL VYAADPAMHAASGNLLLDKRTTCFSVAALTN
19	GPSDASALLYLELAKRLEALEAGKSKSAPKVITKKDAADAKKKMRHKRV ATKAYNPTQAFGLRGP

21	ARLAQMCYAGVTLDVMRINKTHYSNLIGNVPDLFTKYNYALPTNFYGCV HAYYINVTDGRYALATHYPATVITPGGRQPYNSYVATVLTNTV
24	TFSTFEEAALCTFLLNKEMYLKLRSE
26	NVSVSRFNPSTWNRRFGFTEQSVFKPQP
28	IVNVSLVKPTVYVYSRVKN
30	ATIPLQASLPFGWLIGVAFLAVFQSATKIIALNKRWQLALYKGFQFI
31	GETALALLLLDRLNQL
33	VPIGGTTVVRPLVED
35	KINQALHGANLRQDESVYSLY
37	FAYSNRRNFLYIIKLVL
41	LSPRWYFYLLGTGPEA
45	ADLYKLMGHFSWWTAFVT
46	FAGTHITMTTVYHITVSQIQLSLLKVT
48	RVAGSITMRSGSSPAL
49	AAIHYLNSKIVDLAQHFS
51	VYMPASWVMRIMTWLEL
53	ECTYISRFSKISLVS
55	LSVIVQPTHYVRVTFSDPNMWYLR
57	LPKEITVATSRTLSYYKLG
59	ASSLFAQFASVTNVTTAMGKALGR
61	PNNTASWFTALTQHKGKEELRFPR
63	QFPILAEAPTAGAFFFGSRLELAKVQNLSG
64	QLGNSFGAIISSINEIFSRLD
66	KYFKNHTSPDVDLGDIS
69	HYYSDFWFIKIGSRKSARL
71	NLDVIISSIVRQLFK
73	NLDMIISSIVRQLFK
75	SELVIGAVIIRGHLRMAG
77	KIQTAPHISLAMLDIQPE
79	GPMLRLIHLVSLWTF
81	VVPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATE
83	KSAGFPFNKFGKARLYYESLSY
85	ALTTFMITKDSYCKLKNSLSDVAFNRYL
87	RQKRTATKQYNVTQAFGRR
91	GVAFLAVFQSATKIIALNKRWQLAL
93	QNKGLFYTQVYKNMAVYRSLTF
94	QGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDK DPNFKDQVILLNKHIDAYKTFPTE
96	TGDSNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALIWN VKDFMSLSEQLRKQIRSAKKNLFPKLTCAATTRQVVNVVTTKIALKGGK I

98	LFSTVFPLTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHNQDVNLHSS RLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPG NF
100	NDVSFLAHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVFN GVSFSTFEEAALCTFLLNKEMYLKLRSDVLLPFTQYNRYLALYNKYKYF
102	LPSLATVAYFNMVYMPASWWMRIMTWLDMVDTSLKCLKDCVMYASAVL LILMTARTVYDDGARRVWTLNMVLTLYVKVYYGNALDQAISMWALIISVT SN
104	RATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLLKKLKKSLNVA KSEFDRDAAMQRKLEKMAHQAMTQMYKQARSEDKRAKVTSAMQTMFL TML
106	RHINAQVAKSHNIALIWNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCAT TRQVVNVVTTKIALKGGKIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVM
108	EHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDG EVITFDNLKTLSSLREVRTIKVFTTVDNINLHTQVVDMSMTYQQQFGPTY
110	AVVYRGTTTTYKLVNGDYFVLTSTHTVMPLSAPTLVPQEHYVRITGLYPTLN ISDEFSSNVANYQKVGMMQKYSTLQGPPTGKSHFAIGLALYPSARIVYT
113	AIGKIQDLSSTASALGKLQDVVNHNAQALNTLVKQLSSKFGAISSVLNDI LSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKM
114	EVVGNVILKPSDEGVKVTQELGHEDLMAAYVENTSITIKKPNELSLALGLK TIATHG
116	DRYFEIYDGGCLNASEVIVNNLDKSAGHPFNKFGKARVYYESLSYQEQD ELFAMTKRNVLPITITQMNLYAISAKNRARTVAGVSILSTMTNRQYHQKM LKSM
118	YNSDYASKGYIANISAFQQVLYYQNNVFMSEAKCWENDINHGPHEFCS QHTMLVKMDGDDVYLPYPDPSRILGAGCFVDDLLKTDSVLLIERFVSLAI DAY
120	EYKGPVTDVIFYKETSYYTTTIKPVSYKLDGVTYTEIEPKLDGYKKDNAYV ENTSITIKKPNELSLALGLKTIATHGIAAINSVPWSKILAYVKPFLGQAA
124	NARDGCVPLNIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQQ VVDADSKIVQLSEINMDNSQNLAWPLIVTALRANSVVKLQNNELSPVALR
126	PIEFHLDGEVLPLDKLSLLSLREVTKIKVFTTVDNINLHTQLVDMSVTYG QHFGPTY
128	KMAHQAMTQMYKQARSEDKRAKVTSAMQTMFLTMLRKLDNDALNNIIN NARDGCVPLNIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQQ VV
130	VSLIWNVKDYMSLSEQLRKQIRSAAKKNNIPFRLTCATTRQVV
132	LAYMLFTKFFYLLGLSAIMQVFFGYFASHFISN
134	VLKKLKKSLNVAKSEFDRDAAMQRKLEKMAHQAMTQMYKQARSEDKR AKVTSAMQTMFLTML
137	HHANEYRQYLDAYNMMISAGFSLWIYKQFDTYNLWNTFTRLQSLENVAY NVVNKGHFDGH
143	VFISPYNSQNAVASKILGLPTQTVDSSQGSEYDYVIFTQTTETAHSCNVN RFNVAITRAKIGILCIMSRLDYDKLQFTSLEIPRRNVATLQAENVTLGLFK D
161	TKRNVLPITITQMNLYAISAKNRARTVAGVSI
165	GLDSLDSYPALETIQVTISSYKDLT

169	PAYSFLPGVYSVFYLYLTFYFTNDVSFLAHLQWFAMFSPIVPFWIT
176	EKGPFHEFCSQHTMLVKMDGDEVYLPYPDPSRILGAGCFVDDLLKTDSVL LIERFVSLAIDAYPLVHHENPEYQNVFRVYLEYIKKLYNDLGNQILDSYSV
182	RTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGGWHNMLK TVYSDVE
205	QSYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTN GTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSK
209	DVLISSIVRQLFKPL
219	SVLQATLSEFSHLASAELEAAQSSYQKALNSGDASPQVLKALQKAVNIA KNAYEKDKAVARKLERMAEQAMTSMYKQARAEDKKAKIVSAMQTM
229	TSLEIPRRNVATLQAENVTLGLFKDCSKIITGLHPTQAPTHLSVDIKFKTEG LCVDIPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRW
231	GSNLEEEILLMLASLFGTYTWTTLVLSMA
233	LSYFVASFRLFARTRS
241	SSYMAISLKFTTSLSLHKLLQTLVLK

In an embodiment of the invention, the polypeptide may comprise at least 2, 3, 4, 5, 6, 7, 8, 9, 10 or more epitope sequences having the amino acid sequences of any of the SEQ ID Nos shown in Table 1. In a further embodiment of the invention, the polypeptide may comprise a concatemer of epitope sequences having the amino acid sequences of any of the SEQ ID Nos shown in Table 1. The term “concatemer” is used to mean a linear arrangement of repeated sequences that are connected in a continuous chain, with or without spacers as required. The skilled person would understand that the polypeptide of the invention may therefore comprise combinations of epitopes selected from the SEQ ID Nos shown in Table 1 wherein certain epitopes are repeated. For example, the polypeptide may comprise the SEQ ID Nos 5, 5, 5, 8 and 9 in order. Thus, the polypeptide of the invention may comprise a combination of concatenated and non-concatenated sequences. Concatenated polypeptide sequences of the invention may comprise 2, 3, 4, 5, 6, 7, 8, 9, 10 or more repeats of any of the epitopes of the invention, as would be appreciated by the person of skill in the art. This applies to all embodiments of the invention.

Table 1 lists SEQ ID Nos 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 24, 26, 28, 30, 31, 33, 35, 37, 41, 45, 46, 48, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 91, 93, 94, 96, 98, 100, 102, 104, 106, 108, 110, 113, 114,

116, 118, 120, 124, 126, 128, 130, 132, 134, 137, 143, 161, 165, 169, 176, 182, 205, 209, 219, 229, 231, 233 and 241. In Table 1, the sequences for spacer elements and any duplicate hotspot elements are not shown, whereas Figure 5 shows the full list of SEQ ID Nos 1 to 245, including spacers and duplicated
5 hotspots, each with their own SEQ ID No. Figure 5 shows potential groupings (combinations) of any of the sequences shown in Table 1.

For example, SEQ ID No 1 is as follows:

FVDGVPFVVSIGYHYKELGIVMNMVDVDTHTYRLSLKDLLLLYAADP

In Figure 5, SEQ ID No 1 is present in a grouping of SEQ ID Nos 1 to 21. The
10 same sequence is also present as SEQ ID No 67 in a grouping of SEQ ID Nos 22 to 93. The same sequence is also present as SEQ ID No 237 in a grouping of SEQ ID Nos 209 to 245. It will be evident to the skilled person which sequences are duplicated in Figure 5. However, for the avoidance of doubt, duplicated sequences are described in Table 2.

15 Table 2: Listing of duplicated sequences in Figure 5.

SEQ ID No	Duplicated in SEQ ID NOs
1	67, 237
3	89, 225
5	43
7	122, 149, 155, 213
11	22
13	39, 139, 211
17	141, 174
24	188
57	112, 206, 215
66	235
85	167

98	199
100	195
102	201
104	191
106	208
108	197
110	193
113	203
114	172
116	152, 180, 217
118	150, 163, 227
120	145, 184, 239
124	147, 171
126	153
128	135, 178, 221
130	186
132	159, 218
134	190
137	157, 223
143	245
182	243

- It is envisaged that the polypeptide may comprise, in any order, each epitope from any of the following groups of epitope sequences, with or without the spacer sequences: a. SEQ ID No 1 to 21, or b. SEQ ID No 22 to 93, or c. SEQ ID No 94 to 113, or d. SEQ ID No 114 to 134, or e. SEQ ID No 135 to 152, or f. SEQ ID No 153 to 190, or g. SEQ ID No 191 to 208, or h. SEQ ID No 209 to 245. It is also
- 5 envisaged that the polypeptide may comprise each epitope from any of the following groups of epitope sequences arranged in the following numerical order on the polypeptide (i.e. according to SEQ ID no): a. SEQ ID No 1 to 21, or b. SEQ ID No 22 to 93, or c. SEQ ID No 94 to 113, or d. SEQ ID No 114 to 134, or e. SEQ

ID No 135 to 152, or f. SEQ ID No 153 to 190, or g. SEQ ID No 191 to 208, or h. SEQ ID No 209 to 245.

The polypeptide may further comprise one or more spacer regions; preferably wherein when the polypeptide comprises more than one of the epitope sequences
5 of the invention or the variants thereof, the one or more spacer regions are positioned between one or more or each epitope sequence. Alternatively, spacer regions may be positioned between only selected epitope sequences as would be determined according to the specific polypeptide design, as explained above. Therefore, the polypeptide may comprise spacer regions between some epitope
10 sequences and no spacer region between other epitope sequences. In an embodiment of the invention, each of the one or more epitope sequences of the polypeptide have at least 70%, 80%, 90%, 95%, 96%, 97%, 98% or 99% sequence identity to any of the sequences of the invention. In a further embodiment of the invention, the polypeptide is no more than 1400, 1300, 1200,
15 1100, 1000, 900, 800, 700, 600, 500, 400, 300, 200, 150, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10 or 9 amino acids in length.

The term "polypeptide" is used to mean a chemical that is made of two or more amino acids, which combine to make protein/polypeptide. As used herein, the term "polypeptide" denotes the claimed polypeptide, whereas the term "epitope"
20 is used to denote potentially shorter amino acid sequences making up the polypeptide. Though of course, the "epitopes" are also polypeptides as denoted in the art and as would be understood by the skilled person. The term "epitope" as used herein also refers to any part of an antigen that is specifically recognised by receptors on any T cells (T cell receptors), as is its natural meaning. The
25 "epitope" may also refer to any part of an antigen that is specifically recognised by any antibodies or B cells. The term "amino acid" retains its meaning as used in the art and would be understood by the skilled person. The term "variant" is used to refer to amino acid sequences that differ from the indicated SEQ ID No by one or more amino acid residues. This could be a substitution, addition or deletion of
30 one or more amino acids. An example would be the amino acid sequences

ACDEF and ACDEE, whereby the latter is a variant of the former as one amino acid has been substituted. Thus, variants may have amino acids inserted, deleted or substituted from the indicated sequence. It is intended, and would be understood by the skilled person, that the polypeptide of the invention comprises

5 a chain of epitopes, with their sequences selected from any of the sequences of the invention and that the full polypeptide sequence is no longer than 1400 amino acids in length. In an embodiment of the invention, each of the one or more epitope variant sequences of the polypeptide have at least 70%, 80%, 90%, 95%, 96%, 97%, 98% or 99% sequence identity to any of the sequences of the

10 invention. The skilled person would understand that the polypeptide of the invention may therefore comprise combinations of epitopes selected from any of the sequences of the invention having any % identity with the respective SEQ ID No. An example of this would be a polypeptide comprising an epitope with 100% identity to SEQ ID No 1 and 70% identity to SEQ ID No 3. This means that the

15 polypeptide may retain its intended qualities and uses thereof if the amino acid sequence changes. Substitutions, deletions and insertions of amino acids may comprise potential manipulations of the polypeptide of the invention. The polypeptide may also be no more than 1400, 1300, 1200, 1100, 1000, 900, 800, 700, 600, 500, 400, 300, 200, 150, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10 or 9

20 amino acids in length. The varying lengths of the polypeptide are intended to encapsulate the respective beneficial characteristics of having a polypeptide that is shorter, usually termed an "oligopeptide" or one that is longer, termed a "polypeptide". The term "spacer regions" is used to describe additional stretches of amino acids between the epitope sequences. These may be important to avoid

25 additional off-target or autoimmune epitopes being created at the junction between the intended epitopes. The spacer regions of the polypeptide of the invention may be present in instances of either one or multiple of the epitope sequences of the invention, comprising the polypeptide of the invention. The inclusion of spacer regions in the polypeptide of the invention is applicable to all embodiments of the

30 invention. The order of the sequences of the polypeptide may also be important,

and can be tailored to optimise the immunogenicity of the polypeptide, whilst minimising the creation of off-target or autoimmune epitopes.

In an embodiment of the invention, the vaccine composition is an mRNA vaccine. In a further embodiment of the invention, the vaccine composition comprises a
5 pharmaceutically acceptable carrier, diluent, excipient and/or adjuvant. In yet a further embodiment of the invention, the vaccine composition is formulated for parental, oral, sublingual, nasal, naso-oral, or pulmonary administration. In a still further embodiment of the invention, said parental administration is subcutaneous, intradermal, intramuscular, subdermal, intraperitoneal, or intravenous
10 administration.

The term “mRNA vaccine” describes a vaccine that uses a copy of an mRNA molecule to produce an immune response. The vaccine delivers molecules of antigen-encoding mRNA into host immune cells, which used the designed mRNA as a blueprint to build foreign protein, or fragment of a protein, that would normally
15 be produced by a pathogen such as a virus. These protein molecules stimulate an adaptive immune response that teaches the body to identify and destroy the corresponding pathogen. In an embodiment of the invention, the foreign protein produced is of a Betacoronavirus. The mRNA may be delivered by a co-formulation of the mRNA encapsulated in lipid nanoparticles. The mRNA may also
20 be formulated for nasal or intratracheal delivery via a nanoparticle-delivery-based system. This system may comprise a biodegradable poly(amine-co-ester) polymer that forms polyplexes with mRNA so that the vaccine is inhalable.

The term vaccine composition, or vaccine, which from herein may be referred to interchangeably as the “composition”, relates to a biological preparation that
25 induces active acquired immunity to a particular infectious disease, in this case a Betacoronavirus infection. Typically, the vaccine contains an agent, or “foreign” agent, that resembles the infection-causing pathogen, or part of the infection-causing pathogen, which within the prior art has often been a weakened or killed form of said pathogen, or recombinant protein or protein fragment from such

pathogen, or polynucleotide encoding such a protein or protein fragment (Williamson et al. 1995, FEMS Immunology and Medical Microbiology 12 (3-4): 223-230). Such a foreign agent, protein or protein fragment would be recognised by a vaccine-receiver's immune system, which in turn would destroy said agent and develop "memory" against the pathogen, inducing a level of lasting protection against future infections from the same or similar pathogenic sub-species. Through the route of vaccination, including those vaccine compositions of the present invention, it is envisaged that once the vaccinated subject again encounters the same pathogen of which said subject was vaccinated against, the individual's immune system may thereby recognise said pathogen and elicit a more effective defence against infection. A more in-depth description of types of vaccines within the art can be found in US6541003 B1.

The phrase "pharmaceutically acceptable" refers to molecular entities and compositions that do not produce an adverse, allergic or other untoward reaction when administered to a human, as appropriate. The preparation of a pharmaceutical composition that contains the vaccine composition of the present invention will be known to those of skill in the art in light of the present disclosure. Moreover, for human administration, it will be understood that preparations should meet sterility, pyrogenicity, general safety and purity standards. A specific example of a pharmacologically acceptable carrier as described herein is borate buffer or sterile saline solution (0.9% NaCl).

As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, surfactants, antioxidants, preservatives {e.g., antibacterial agents, antifungal agents}, isotonic agents, absorption delaying agents, salts, preservatives, drugs, drug stabilizers, gels, binders, excipients, disintegration agents, lubricants, sweetening agents, flavouring agents, dyes, such like materials and combinations thereof, as would be known to one of ordinary skill in the art (see, for example, Remington's Pharmaceutical Sciences, 18th Ed. Mack Printing Company, 1990, pp. 1289-1329).

- Examples of adjuvants which may be effective include but are not limited to: unmethylated cytosine-guanine dinucleotide (CpG) motifs, granulocyte-macrophage colony-stimulating factor (GM-CSF), aluminium hydroxide, N-acetylmuramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-
- 5 isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80
- 10 emulsion. Further examples of adjuvants and other agents include aluminium hydroxide, aluminium phosphate, aluminium potassium sulfate (alum), beryllium sulfate, silica, kaolin, carbon, water-in-oil emulsions, oil-in-water emulsions, muramyl dipeptide, bacterial endotoxin, lipid X, *Corynebacterium parvum* (*Propionobacterium acnes*), *Bordetella pertussis*, polyribonucleotides, sodium
- 15 alginate, lanolin, lysolecithin, vitamin A, saponin, liposomes, levamisole, DEAB-dextran, blocked copolymers or other synthetic adjuvants. Such adjuvants are available commercially from various sources, for example, Merck Adjuvant 65 (Merck and Company, Inc., Rahway, N.J.) or Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, Mich.).
- 20 Thus, in some embodiments of the invention the composition may further comprise a pharmaceutically acceptable carrier, diluent, excipient and/or adjuvant. In a preferred embodiment, the composition may further comprise an adjuvant.
- In a second aspect of the invention, there is provided a vaccine composition according to the first aspect of the invention, for therapeutic or prophylactic use.
- 25 In a third aspect of the invention, there is provided a vaccine composition according to the first aspect of the invention, for use in the treatment or prophylaxis of an infection caused by a Betacoronavirus.

In a fourth aspect of the invention, there is provided a use of a vaccine composition according to the first aspect of the invention, in the manufacture of a medicament for the treatment of a Betacoronavirus infection.

5 In a fifth aspect of the invention, there is provided a method of treating a Betacoronavirus infection, comprising administering to a subject an effective amount of the vaccine composition according to the first aspect of the invention.

10 In a sixth aspect of the invention, there is provided a vector encoding the polypeptide as defined in the first aspect of the invention, preferably wherein the vector may further comprise regulatory elements capable of driving transcription and/or translation of the polynucleotide in a host cell. The vector may also be a viral vector. By "vector" we intend any vehicle that is capable of supporting (i.e. carrying, encapsulating, incorporating and/or protecting) the polynucleotide of the invention and facilitating stable or transient transfection of a target host cell with
15 the polynucleotide of the invention. The vector may also include the indicated regulatory elements to drive expression of a polypeptide encoded by the polynucleotide in the host cell. Suitable vectors may be, for example, plasmids, viral particles (e.g. lentiviruses, adenoviruses, AAVs and any number of other viruses, as would be understood by a person of skill in the art), nanoparticles, lipid-nanoparticles etc. An example of a vector suitable for the invention is a lentivirus, capable of inserting DNA into host cell genomes. Examples may also include one or more adenovirus vectors, vesicular stomatitis virus vectors, influenza virus vectors or measles virus vectors. When viral particles are used as vectors, they are typically modified from their wild-type form to remove viral antigen encoding
20 nucleic acid and/or to prevent viral replication in the host cell. The vector may be a naked DNA such as a plasmid or a virus, among other vectors used in the art and that would be familiar to the skilled person. The term "regulatory elements" refers to nucleotide sequences of genes that are involved in regulation of transcription.

In a seventh aspect of the invention, there is provided a microorganism comprising a polypeptide as defined in the first aspect of the invention, or a polynucleotide encoding the polypeptide as defined in the first aspect of the invention. The
5 microorganism may be a bacterium. The term "microorganism" refers to a microscopic organism, such as a bacterium, virus or fungus. Microorganisms may be used to further propagate the polynucleotide or polypeptide of the invention or to be administered as part of therapeutic or prophylactic use in order to deliver the polynucleotide or polypeptide to a specific tissue or cell type such as an antigen
10 presenting cell. A microorganism comprising a polypeptide of the invention may be suited to industrial production of the polypeptide, and may include: *Escherichia coli*, *Bacillus subtilis*, *Streptomyces spp.*, *Corynebacterium glutamicum*, *Pseudomonas putida*, *Clostridium spp.* and *Lactobacillus*. The microorganism of the invention may also be genetically modified. The polypeptide of the invention
15 may also be produced by synthetic methods, of which the skilled person would be aware. The term "prophylactic treatment", as used herein, refers to a medical procedure whose purpose is to prevent or reduce the morbidity or duration of (rather than treat or cure) a disease, such as an infection. In contrast, the term "therapeutic treatment" refers to a medical procedure with the purpose of treating
20 or curing a viral infection or the associated symptoms thereof, as would be appreciated within the art.

In an eighth aspect of the invention, there is provided a polynucleotide as defined in the first aspect of the invention, or a vector according to the fifth aspect of the invention, formulated in a nanoparticle or lipid formulation. The term
25 "nanoparticle" refers to a nanoscale particle between 1 and 100 nanometres in diameter, often with a very high surface area to volume ratio. Formulation in a nanoparticle may improve delivery of the polynucleotide or vector of the invention to a subject and/or the immune system of a subject. The nanoparticle may be a lipid nanoparticle, designed to facilitate encapsulation and delivering of the
30 polynucleotide or vector of the invention.

In a ninth aspect of the invention, there is provided a composition comprising: i) one or more of the epitope sequences comprising any one or more of SEQ ID Nos 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 24, 26, 28, 30, 31, 33, 35, 37, 41, 45, 46, 48, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 91, 93, 94, 96, 98, 100, 102, 104, 106, 108, 110, 113, 114, 116, 118, 120, 124, 126, 128, 130, 132, 134, 137, 143, 161, 165, 169, 176, 182, 205, 209, 219, 229, 231, 233 and/or 241, or a variant thereof having at least 70% sequence identity, each of said epitope sequences or variants thereof being no more than 1400 amino acids in length, or ii) one or more polynucleotides encoding one or more epitope sequences comprising any one or more of SEQ ID Nos 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 24, 26, 28, 30, 31, 33, 35, 37, 41, 45, 46, 48, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 91, 93, 94, 96, 98, 100, 102, 104, 106, 108, 110, 113, 114, 116, 118, 120, 124, 126, 128, 130, 132, 134, 137, 143, 161, 165, 169, 176, 182, 205, 209, 219, 229, 231, 233 and/or 241, or a variant thereof having at least 70% sequence identity, each of said epitope sequences or variants thereof being no more than 1400 amino acids in length.

In a tenth aspect of the invention, there is provided a vector, microorganism or composition according to the above aspects of the invention for therapeutic or prophylactic use. The vector, microorganism or composition according to the above aspects of the invention may be for use in medicine.

In an eleventh aspect of the invention, there is provided a vector, microorganism or composition according to the above aspects of the invention, for use in the treatment or prophylaxis of an infection caused by Betacoronavirus. The terms "treatment" and "prophylaxis" are to be used interchangeably with "therapeutic treatment" and "prophylactic treatment" respectively. It is envisaged that the polypeptide, polynucleotide, vector, microorganism or composition of the present invention may be used against any Betacoronavirus infection.

Betacoronaviruses, from the family *Coronaviridae*, are a group of enveloped, positive-sense single-stranded RNA ((+ssRNA) viruses which can cause respiratory tract infections in human hosts. Mild Betacoronavirus infections include some cases of the common cold, whilst more lethal species of

5 Betacoronavirus such as severe acute respiratory syndrome-related coronavirus (SARS-CoV-1), Middle East respiratory syndrome-related coronavirus (MERS-CoV), and severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2), can cause the more serious diseases SARS, MERS, and COVID-19, respectively. It is proposed that SARS-CoV-2 shares zoonotic origins and close

10 genetic similarity with SARS-CoV-1, and as such much of our understanding of COVID-19, as well as the research and development of potential prophylactic and therapeutic treatments, has come from the analysis of such other coronaviruses.

SARS-CoV-2 is the causative viral agent behind the 2019-2020 pandemic of COVID-19, a respiratory syndrome characterised by high fever, malaise, rigors,

15 headache, dry cough, lymphopenia and progression to interstitial infiltration in lungs with an eventual mortality of greater than 10% in many countries. SARS-related pathologies of the lungs involve the subsequent stages of viral replication, immune system hyperactivation, and pulmonary destruction (Weis & Navas-Martin 2005, Microbiol Mol Biol Rev. 69 (4): 635-64) and inflammatory exudates

20 in the lungs.

Betacoronaviruses, such as SARS-CoV-2, attach to their specific cellular receptors via the viral spike protein-invading cells lining the respiratory tract. The receptor for the SARS-CoV-2 virus, a positive single stranded RNA ((+)ssRNA) coronavirus, was identified as angiotensin-converting enzyme 2 (ACE2): a zinc

25 metalloprotease (Li et al. 2003, Nature 426: 450-454). Diseased lungs show diffuse alveolar damage, epithelial cell proliferation, and an increased number of macrophages. Further, multinucleate giant-cell infiltrates of macrophage or epithelial cells with syncytium-like cell formation have been described. In addition to hemophagocytosis in the lung, lymphopenia and white-pulp atrophy of the

30 spleen have been observed in SARS patients. At present, most COVID-19

patients receive traditional supportive care such as breathing assistance and/or steroid therapy.

It is envisaged that the polypeptide, polynucleotide, vector, microorganism or composition of the present invention may aid in the therapeutic or prophylactic treatment of a Betacoronavirus infection.

In a twelfth aspect of the invention, there is provided a vector, microorganism or composition for use according to the eleventh aspect of the invention, wherein the infection is caused by Embecovirus, Hibecovirus, Merbecovirus, Nobecovirus, Sarbecovirus, MERS-CoV, SARS-CoV or SARS-CoV-2. It is envisaged that the virus species may comprise Betacoronavirus 1, Bovine coronavirus, Human coronavirus OC43, China Rattus coronavirus HKU24, Human coronavirus HKU1, Murine coronavirus, Myodes coronavirus 2JL14, SARS-CoV-1, SARS-CoV-2, Bat SARS-like coronavirus WIV1, Bat coronavirus RaTG13, Hedgehog coronavirus 1, MERS-CoV, Pipistrellus bat coronavirus HKU5, Tylonycteris bat coronavirus HKU4, Eidolon bat coronavirus C704, Rousettus bat coronavirus GCCDC1, Rousettus bat coronavirus HKU9, Bat Hp-betacoronavirus Zhejiang2013 or any other bat coronavirus. The skilled person would understand that this is not an exhaustive list of virus species and that any embodiment that a "Betacoronavirus" is referred to may comprise one of the above virus species.

20

The invention also provides for the use of the polypeptide, polynucleotide, vector, microorganism, composition or vaccine composition of the invention in the manufacture of a medicament for the treatment of a Betacoronavirus infection.

25 The present invention provides a method of treating a Betacoronavirus infection, comprising administering to a subject an effective amount of the polypeptide, polynucleotide, vector, microorganism, composition or vaccine composition of the present invention.

It is envisaged that the polypeptide, polynucleotide, vector, microorganism, composition or vaccine composition of the present invention may aid in the therapeutic or prophylactic treatment of a Betacoronavirus infection in a human subject, wherein said composition comprises one or more epitopes of the present invention that are capable of stimulating a broad adaptive immune response across a variety of HLA types. The term "subject" refers to a person subjected to treatment, observation or experiment and would be understood by the skilled person as such.

The active acquired immunity that may be induced by the invention is expected to be predominantly cellular. The active acquired immunity that may be induced by the invention may also be humoral. Humoral immunity refers to a response involving B cells which produce antibodies that specifically bind to antigens, or any future antigens, corresponding to those within the administered vaccine composition. B cells, each expressing a unique B cell receptor (BCR), recognise antigens in their native form. Upon this recognition and further interaction with other cells of the immune system, the activated B cell can differentiate into a plasma cell specialised to secrete antibodies against the encountered antigen. The term antibody refers to an immunoglobulin (Ig) that is produced by the immune system to specifically identify and neutralise foreign antigens. A subset of these B-cell derived plasma cells become long-lived antigen-specific memory B cells, as would be well understood by the skilled person.

Cellular immunity, meanwhile, can be broken into two distinct arms. The first involves helper T cells, or CD4+ T cells, which produce cytokines and orchestrate the activity of other immune cells in the immune response. The second involves killer T cells, also known as cytotoxic T lymphocytes (CTLs), or CD8+ T cells, which are cells capable of recognising antigens/epitopes presented by HLA and eradicate viral or bacterial infected host cells or cancerous or otherwise diseased host cells. In contrast to B cells, T cells only recognise antigens that have been processed into peptides and have been loaded onto histocompatibility complex (MHC) molecule and presented at the cell surface. CD4+ T cells interact with

MHC class II molecules (MHC Class II), and are responsible for orchestrating the immune response, recognizing foreign antigens, activating various parts of the immune system and activating B cells and CD8+ T cells. CD8+ T cells interact with MHC Class I receptors and play a role in mounting an immune response against intracellular pathogens. As would be understood by the skilled person, on resolution of the infection, a subset of both CD8+ T cells and CD4+ T cells may remain as memory T cells, contributing to the acquired adaptive immunity, and allowing for a faster and stronger response to any secondary infection from the same foreign body (Bonilla & Oettgen 2010, Journal of Allergy and Clinical Immunology 125: 33-40).

It is envisaged that the vaccine composition of the present invention may be an epitope-based vaccine, or in other words, is comprised of one or more epitopes. Epitope-based vaccines (EVs) make use of short antigen-derived peptides corresponding to immune epitopes, which are administered to trigger a protective cellular immune response. EVs may also potentially enhance humoral responses. EVs potentially allow for precise control over the immune response activation by focusing on the most relevant — immunogenic and conserved — antigenic regions. Experimental screening of large sets of peptides is time-consuming and costly; therefore, *in silico* methods that facilitate T-cell epitope mapping of protein antigens are paramount for EV development. The prediction of T-cell epitopes focuses on the presentation of peptides at the infected cell surface by proteins encoded by the major histocompatibility complex (MHC).

The epitopes of the present invention may interact with MHC Class I and/or MHC Class II molecules to induce a CD8+ T cell and/or CD4+ T cell response, respectively. In a preferred embodiment of the present invention, there may be at least one epitope that interacts with MHC Class I, and at least one epitope that interacts with MHC Class II.

An “antigen” refers to a molecule capable of being bound by a receptor on a T cell, and may be comprised of one or more epitopes. An “antigen” may also refer to a

molecule capable of being bound by an antibody or a B cell. As such, the terms epitope and antigen may be used interchangeably herein. Epitopes may also be referred to by the molecule for which they bind, such as "T cell epitopes", or more specifically, "MHC Class I epitopes" or "MHC Class II epitopes".

- 5 Thus in some embodiments of the present invention, the one or more epitopes may have the same length, or same number of amino acids. In other embodiments, the one or more epitopes may differ in length, or the number of amino acids. In some embodiments, the one or more epitopes may overlap with each other at least partly. In other embodiments, the one or more epitopes may
10 overlap across more than one hotspot.

- In other embodiments, one of the epitopes may fully comprise the entirety of another epitope within the same composition. It is envisaged that the one or more epitopes of the present invention are capable of stimulating a broad adaptive immune response across a plurality of human leukocyte antigen (HLA) types. In
15 the context of the present invention, the term "plurality" is used to refer to "at least two", or "two or more". The human leukocyte antigen (HLA) complex system is a set of genes encoding the MHC proteins in humans. Owing to the highly polymorphic nature of HLA genes, in which the term "polymorphic" refers to a high variability of different alleles, the precise MHC proteins of each human individual
20 coded by varying HLA genes may differ to fine-tune the adaptive immune system. Many thousands of different alleles have been recognised for HLA molecules. As a result, each individual may have a unique "HLA type", or HLA genotype, that differs across the global population, with a slight variability in the functioning of the immune system. The terms HLA type, HLA allele, or HLA genotype may be used
25 interchangeably herein. HLA types are of particular significance when considering a vaccine comprised of epitopes that interact with MHC class I or class II molecules, as many epitopes are restricted in their capability of binding only particular HLA molecules encoded by particular HLA alleles, or in other words, are restricted to certain HLA types only. It would thus be appreciated by the skilled
30 person that T cell epitopes that are capable of binding to a subject's MHC Class I

or MHC Class II molecules (and be presented at the infect cell surface), compatible with said subject's HLA type, would thus present as a robust vaccine. A vaccine composition consisting of the same T cell epitopes may not prove effective if given to a subject with a different HLA type, if said HLA type encodes

5 MHC molecules that are not capable of binding or interacting with said T cell epitopes. Such epitopes would not be able to stimulate a broad adaptive immune response across for either MHC Class I and/or MHC Class II immunogenicity in that particular subject.

The polypeptides, polynucleotides, vectors, microorganisms, compositions,

10 vaccine compositions and epitopes of the present invention, in contrast, are envisaged to be able to stimulate a broad adaptive immune response across a plurality of HLA types, including alleles such as HLA-A*24:02 and HLA-B*07:01. The HLA alleles as referenced herein are given contemporary HLA nomenclature as standard to the field, wherein HLA-A, for example, refers to the gene loci in

15 chromosome 6, whilst HLA-A*24:02 refers to the protein the allele codes for. An in-depth explanation of the complexities of HLA nomenclature can be found in Marsh *et al.* 2010, Tissue Antigens 75(4): 291-455. The approach of the present invention analysed 150 of the most frequent HLA molecules in the human population, as shown in Figures 1-2.

20 The platform used to identify and predict the one or more epitopes of the present invention was surprisingly robust, as was its integrated statistical analysis. Firstly, epitope mapping of the proteome for Betacoronavirus for Class I epitopes was carried out using cell-surface antigen presentation and immunogenicity predictors from the "NEC Immune Profiler" suite of tools

25 (<https://pubmed.ncbi.nlm.nih.gov/33361777/>). Antigen Presentation (AP) was predicted from a machine learning model that integrates in an ensemble machine learning layer information from several HLA binding predictors - trained using empirically measured binding affinity data - and 13 different predictors of antigen processing. The AP scores were calculated based on a set of 150 most frequent

30 Class I HLA alleles (A and B) for all the possible combinations between an HLA

allele and each epitope. As such, each amino acid within a considered protein sequence was assigned an AP score. Matrices with the AP scores at each amino acid position for each of the considered HLA alleles were generated for each protein and genotype. Hotspots were identified as protein regions consistently
5 showing a higher-than-average AP score within each matrix, across all considered HLA alleles.

Firstly, the most prevalent amino acid sequence for each protein and genotype was identified within the input data set. Next, the selected amino acid sequences of the proteins were split into peptides of length 9 and 10 by means of a sliding
10 window. The important determinants of antigen presentation (AP) were assessed for each peptide of each protein and each genotype for their potential to be efficiently presented. These determinants consisted of: (1) the predicted binding affinity between the candidate peptide and 150 of the most frequent Class I HLA molecules in the human population, (2) the predicted potential of the candidate
15 peptide to be efficiently processed by the antigen processing machinery of the host infected cell, and (3) the predicted probability of the candidate peptide to be presented on the host infected cell surface.

To assess the performance of the identified hotspots, "Digital Twin" (DT) simulations based on statistical models of optimized population coverage were
20 performed. These analyses estimated the response likelihood of the hotspots against different HLA haplotypes of individuals from different regions of the world, as well as on a worldwide population. For this purpose, the HLA haplotype data from The Allele Frequency Net Database was leveraged
25 (<https://academic.oup.com/nar/article/48/D1/D783/5624967>).

Top candidate hotspots were selected as being the best performing with respect to their estimated population coverage (i.e., response likelihoods) from DT simulations and being the most dissimilar with respect to their amino acid

sequence (e.g., differing by more than 90% in sequence). Finally, the homologs of each of the top hotspots in other genotypes were identified.

This approach advantageously uses a statistical model to quantitatively analyse the predicted immunogenic potential of one or more epitopes – in other words the predicted ability of the one or more epitopes to instigate an immunogenic response – within an amino acid sub-sequence, across a set of different HLA types. The candidate regions (or “hotspots”) of the amino acid sequence that are identified by the quantitative statistical analysis may represent regions (or areas) of the one or more source proteins that are most likely to be viable vaccine targets and may be used in vaccine design and creation. These source proteins include those taken from species of Embecovirus, Hibecovirus, Merbecovirus, Nobecovirus, Sarbecovirus, MERS-CoV, SARS-CoV or SARS-CoV-2.

It is envisaged that each of the hotspots identified herein may comprise one or more epitopes capable of stimulating an adaptive immune response through MHC Class I and/or MHC Class II. A candidate region may comprise a single epitope that is predicted to instigate an immunogenic response across a plurality of the HLA types. Such an epitope may be termed as “overlapping with” a number of HLA types. More typically however, a candidate region comprises a plurality of epitopes that, collectively, overlap with a large proportion of the analysed HLA types. For example, one epitope within a candidate region may overlap with n HLA types and a different epitope within the candidate region may overlap with m HLA types such that the candidate region is predicted to instigate an immunogenic response across the $(m+n)$ HLA types.

The approach comprised the step of assigning, for each of the set of HLA types, an antigen presentation (AP) score for each amino acid, wherein said score is indicative of the immunogenic potential of an epitope comprising that amino acid, for that HLA type. For a given HLA allele, the score allocated to an amino acid corresponds to the best score obtained by an epitope prediction overlapping with this amino acid. For Class I HLA alleles, 1 represents the best score, wherein the

- amino acid has a higher likelihood of being naturally presented on the cell surface, whereas a score closer to 0 represents a lower likelihood. For Class II HLA alleles, in contrast, the predictions are of percentile rank binding affinity scores wherein lower scores are best. With a range of possible output scores of 0 to 100 for Class
- 5 II HLA alleles, a score of 0 represents the best score, with the highest binding affinity. As explained above, AP scores for a hotspot can be combined and represented by the sum of their values. A given hotspot may contain multiple epitopes, and as such will have an AP score that reflects this. For example, an AP score of ~60 may indicate that a given hotspot contains many predicted
- 10 epitopes, such that the hotspot is a desirable target for a potential vaccine. Furthermore, a hotspot that has a low AP score may still be desirable, as this may indicate the presence of a unique epitope sequence for a particular HLA type. This would be indicated by an AP score between 0 and 1, for a given hotspot. A hotspot that contains a low number of immunogenic epitopes for a low number of HLA
- 15 types may have a summed AP score of greater than 1 but below 10. While hotspots such as these may not be strongly immunogenic, their use to provoke immune responses in individuals with HLA types that would otherwise be difficult to target may be desirable. The summed AP scores for the identified hotspot regions appear in Table 3.
- 20 The predictions for Class I and Class II HLA types were performed using an antigen presentation and binding affinity prediction algorithm, as well as experimental data. Examples of publicly available databases and tools that may be used for such predictions include the Immune Epitope Database (IEDB) (<https://www.iedb.org/>), the NetMHC family of prediction tools
- 25 (<http://www.cbs.dtu.dk/services/NetMHC/>), the TepiTool prediction tool (<http://tools.iedb.org/tepitool/>), the NetChop prediction tool (<http://www.cbs.dtu.dk/services/NetChop/>) and the MHC-NP prediction tool (<http://tools.immuneepitope.org/mhcnpl/>). Other techniques are disclosed in WO2020/070307 and WO2017/186959.

Antigen presentation was predicted from a machine learning model that integrates in an ensemble machine learning layer information from several HLA binding predictors and a plurality of different predictors of antigen processing (trained on mass spectrometry data).

- 5 Each of the identified epitopes was then preferably allocated a score based on the immunogenic potential predicted using the above techniques. Advantageously, the method not only identified candidate regions comprising epitopes that may bind to an HLA molecule, but also those CD8 epitopes that are naturally processed by a cell's antigen processing machinery, and presented on the surface of host
10 infected cells.

The AP scores were assigned by the following protocol. Firstly, a plurality of epitopes were identified across the amino acid sequence, in a "moving window" of amino acids of fixed length. This was performed for each HLA type. For each of the identified first epitopes, a score was generated that is indicative of the
15 immunogenic potential of that epitope, for the respective HLA type. A plurality of further epitopes were subsequently identified across the amino acid sequence, for each HLA type. Again, this was performed using a "moving window approach". Each of the further epitopes were also assigned a score that was indicative of the immunogenic potential of that epitope, for the respective HLA type. Each amino
20 acid was then assigned, for each HLA type, the score of the epitope that was predicted to have the best immunogenic potential of all the epitopes comprising that amino acid. Hence, for a particular HLA type, if epitope "A" and epitope "B" both comprised a particular amino acid "X", the amino acid "X" would have been assigned the score of whichever epitope "A" or "B" is predicted to have the best
25 immunogenic potential. In other words, for a given HLA type, the score allocated to an amino acid corresponds to the best score obtained by an epitope overlapping with this amino acid.

The AP score for each amino acid within a given source protein, or open reading frame, was averaged across HLA types. An AP score is given for each amino

acids, wherein it is the average AP score of that amino acid across 150 of the most common HLA-A and HLA-B alleles that correspond to MHC Class I. In total, 150 of the most common human HLA-A and HLA-B across the globe were subjected to analysis. The one or more epitopes of the present invention may, further to
5 being able to interact with the top 150 HLA-A or HLA-B alleles, also be able to stimulate a broad adaptive immune response across a plurality of HLA-C genotypes and class II genotypes of the HLA-DP, HLA-DQ and/or HLA-DR alleles.

The HLA types analysed may further be characterised into HLA types of the same or different human population groups. A population group may be an ethnic
10 population group (e.g. Caucasian, Africa, Asian) or a geographical population group (e.g. Lombardy, Lagos).

It is envisaged that the polypeptide, polynucleotide, vector, microorganism, composition or vaccine composition of the present invention comprises one or more epitopes found within any one or more of the hotspots, including the
15 sequences of any of SEQ ID NOs: 1-245, and wherein said epitopes meet a particular threshold of a mean antigen presentation (AP) cut off value. Said mean AP cut off value is the value, averaged across all amino acids within an epitope, for which said epitope is considered able to stimulate a broad adaptive immune response across a plurality of HLA types, for either MHC Class I and/or MHC Class
20 II immunogenicity. There is a tradeoff between the immunogenicity of the polypeptide, polynucleotide, vector, microorganism, composition or vaccine composition of the present invention and the number of Betacoronavirus subgenera they are cross-reactive to, as seen in Figure 3. An exemplary AP cut off value is 0.65 for at least 50 HLA types. This would identify epitopes that are
25 broadly protective in terms of the number of Betacoronavirus subgenera they are immunogenic towards and the number of HLA types they would instigate an immune response in. Further exemplary AP cut off values include 0.5, 0.6 and 0.7 for any number of HLA types, up to 150.

For the sake of avoiding confusion, the term “antigen presentation (AP) value” may be used to mean binding affinity or percentile ranking, and the terms shall be used interchangeably. As such, reference to a mean “AP cut off value” in the context of MHC Class II, is to be construed as the mean binding affinity or mean
5 percentile ranking of the relevant epitopes.

There is also a tradeoff between the Betacoronavirus species coverage of the polypeptide, polynucleotide, vector, microorganism, composition or vaccine composition of the present invention and the human population coverage they are cross-reactive to by HLA-A frequency, as seen in Figure 4.

10 It is envisaged that the polypeptide, polynucleotide, vector, microorganism, composition or vaccine composition of the present invention may comprise any number of epitopes as would be suitable for use. In some embodiments, the composition comprises at least 5 epitopes. In a preferred embodiment, the composition comprises between 5 and 10 epitopes. In a yet further preferred
15 embodiment, the composition comprises between 5 and 20 epitopes.

It is envisaged that the composition of the present invention may comprise an immunogenic portion of the Betacoronavirus, wherein the term “immunogenic portion” refers to one or more epitopes found within any one or more of Table 1 or Figure 5, or a polynucleotide encoding said epitope. Each epitope considered
20 must be able to stimulate a broad adaptive immune response across a plurality of HLA types, for either MHC Class I and/or MHC Class II immunogenicity.

In some embodiments, the polypeptide of the invention may comprise at least 2 epitope sequences having the amino acid sequences of the invention. In another embodiment, the polypeptide of the invention may comprise at least 3 epitope
25 sequences having the amino acid sequences of the invention. In another embodiment, the polypeptide of the invention may comprise at least 4 epitope sequences having the amino acid sequences of the invention. In another embodiment, the polypeptide of the invention may comprise at least 5 epitope

sequences having the amino acid sequences of the invention . In another embodiment, the polypeptide of the invention may comprise at least 6 epitope sequences having the amino acid sequences of the invention . In another embodiment, the polypeptide of the invention may comprise at least 7 epitope
5 sequences having the amino acid sequences of the invention . In another embodiment, the polypeptide of the invention may comprise at least 8 epitope sequences having the amino acid sequences of the invention . In another embodiment, the polypeptide of the invention may comprise at least 9 epitope sequences having the amino acid sequences of the invention . In another
10 embodiment, the polypeptide of the invention may comprise at least 10 epitope sequences having the amino acid sequences of the invention .

In some embodiments, the size of said immunogenic portion may have an upper limit of 1400 amino acids in length. In another embodiment, the upper limit may be 1300 amino acids in length. In another embodiment, the upper limit may be
15 1200 amino acids in length. In another embodiment, the upper limit may be 1100 amino acids in length. In another embodiment, the upper limit may be 1000 amino acids in length. In another embodiment, the upper limit may be 900 amino acids in length. In another embodiment, the upper limit may be 800 amino acids in length. In another embodiment, the upper limit may be 700 amino acids in length.
20 In another embodiment, the upper limit may be 600 amino acids in length. In another embodiment, the upper limit may be 500 amino acids in length. In another embodiment, the upper limit may be 400 amino acids in length. In another embodiment, the upper limit may be 300 amino acids in length. In another embodiment, the upper limit may be 200 amino acids in length. In another
25 embodiment, the upper limit may be 150 amino acids in length. In another embodiment, the upper limit may be 100 amino acids in length. In another embodiment, the upper limit may be 90 amino acids in length. In another embodiment, the upper limit may be 80 amino acids in length. In another embodiment, the upper limit may be 70 amino acids in length. In another
30 embodiment, the upper limit may be 60 amino acids in length. In another

embodiment, the upper limit may be 50 amino acids in length. In other embodiments, the upper limit may be 40 amino acids in length. In a further embodiment, the upper limit may be 30 amino acids. In yet another further embodiment, the upper limit may be 20 amino acids in length. In yet another
5 further embodiment, the upper limit may be 10 amino acids in length. In yet another further embodiment, the upper limit may be 9 amino acids in length. Accordingly, the immunogenic portion may consist of the complete sequence defined herein as a hotspot, or fragments thereof that comprise at least one of the epitopes defined herein.

10 It is envisaged that such an immunogenic portion for use in the composition of the present invention would be recombinant in nature, wherein recombinant refers to the artificial and/or modified characteristic of said immunogenic portion, which may be produced through genetic recombination means. As such, it is envisaged that the immunogenic portion may be a non-functional, recombinant fragment of
15 a protein, wherein said non-functional, recombinant fragments include one or more of the epitopes capable of stimulating a broad adaptive immune response across a plurality of HLA types, as described in the present invention.

The epitope sequences of the invention are also considered to be able to stimulate a broad adaptive immune response across a plurality of HLA types, for either MHC
20 Class I and/or MHC Class II immunogenicity.

In some embodiments, the invention may comprise one or more epitopes according to the present invention that are considered able to stimulate a broad adaptive immune response across a plurality of HLA types for MHC Class I. In other embodiments, the invention comprises one or more epitopes according to
25 the present invention that are considered able to stimulate a broad adaptive immune response across a plurality of HLA types for MHC Class II. In a preferred embodiment, the invention may comprise one or more epitopes that are considered able to stimulate a broad adaptive immune response across a plurality of HLA types for both MHC Class I and MHC Class II.

It is envisaged that the present invention may further comprise tertiary protein structures, or domains thereof, of species including members of the Embecovirus, Hibecovirus, Merbecovirus, Nobecovirus, Sarbecovirus subgenera, or by MERS-CoV, SARS-CoV or SARS-CoV-2.

- 5 In some embodiments of the present invention, the vaccine composition may further comprise a pharmaceutically acceptable carrier, diluent, excipient and/or adjuvant, as well as minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine.
- 10 In some embodiments, the polypeptide, polynucleotide, vector, microorganism, composition or vaccine composition may be used in the therapeutic or prophylactic treatment of any Betacoronavirus infection. In a preferred embodiment, the Betacoronavirus infection may be caused by members of the Embecovirus, Hibecovirus, Merbecovirus, Nobecovirus, Sarbecovirus subgenera, or by MERS-
- 15 CoV, SARS-CoV or SARS-CoV-2.

- The one or more compositions of the present invention may be formulated for parenteral, oral, sublingual, nasal, naso-oral, or pulmonary administration. In a preferred embodiment, the parenteral administration may be subcutaneous,
- 20 intradermal, intramuscular, subdermal, intraperitoneal or intravenous.

- It is envisaged that administration of the polypeptide, polynucleotide, vector, microorganism, composition or vaccine composition according to the present invention would be carried out following an appropriate immunisation regimen. The term "appropriate immunisation regimen" is to be construed as a schedule or
- 25 timescale of one or more administrations of the compositions of the present invention, which may resultantly yield the most effective results in consideration of immunisation efficacy and safety of the subject to which the composition is being administered. For example, for the therapeutic or prophylactic treatment of

a Betacoronavirus infection, an immunisation regimen should be chosen that yields as effective immunisation against the Betacoronavirus as possible, whilst still maintaining suitable safety for the subject.

In some embodiments of the present invention, the immunisation regimen may
5 comprise a single administration. In other embodiments, the immunisation regimen may comprise multiple administrations, either concomitantly or over an appropriate period of time.

It is envisaged that the appropriate dosage regimen may be repeated for a subject at a suitable time.

10 There exists further the possibility to further administer boost immunisations after a more extended period of time. This may be selected as an appropriate measure if a subject's T-cell response falls below determined protective levels. The boost immunisations may be administered if a subject's immunoglobulin G (IgG) antibody levels fall below determine protective levels. Thus in some
15 embodiments, an appropriate dosage regimen may be given as a "boost immunisation" after 6 months.

In some embodiments of the present invention, the polypeptide, polynucleotide, vector, microorganism, composition or vaccine composition may be administered for the treatment or prevention of infections caused by a virus in combination with
20 one or more other antiviral therapies or other appropriate therapies such as stem cell therapies. Such antiviral therapies may include administration of oseltamivir phosphate (Tamiflu®), zanamivir (Relenza®), peramivir (Rapivab®), baloxavir marboxil (Xofluza®), or lopinavir/ritonavir (Aluvia®). Such antiviral therapies may be administered simultaneously, separately or sequentially with the composition
25 of the present invention. In a further embodiment, the antiviral therapy is administered via the same or different route of administration as the composition of the present invention, for example via intradermal injection.

Example 1

The first part of the data processing to identify potential epitopes involved the generation of epitope scores for each amino acid position in all the proteins in the Betacoronavirus proteome, for 150 HLA types.

- 5 For HLA types of MHC class I, the scores assigned to each amino-acid were in the range of 0 to 1, with 1 being the best epitope score. For HLA types of MHC class II, the scores assigned to each amino-acid were in the range of 0 to 100 (percentile ranks), with 0 being the best epitope score. A score for a designated amino-acid was determined as the best score that a peptide overlapping that
- 10 amino-acid carries in the predictions. All peptides of size 8-12 for class I, and size 15 for class II had been processed by the antigen presentation framework. At this point, one dataset per protein was generated. Each row in the dataset represented the amino-acid epitope scores predicted for one HLA type.

Example 2

- 15 The sequences of the peptides were then individually run through the inventor's AI prediction platform, the NEC Immune Profiler (NIP), which provided the AI predictions for these key determinants, such as the AP scores surface. The predictions were run individually for each protein and genotype of interest. The AP score is in the range between 0 and 1, with 1 being the maximum of the
- 20 likelihood that a specific candidate peptide was presented on the host infected cell. In brief, the amino acid sequences of the proteins were split into minimal epitopes (9- and 10-mers) by means of a sliding window. The important determinants of antigen presentation (AP) were assessed for each peptide of each protein and each genotype for their potential to be efficiently presented. These
- 25 determinants consisted of: (i) the predicted binding affinity between the candidate peptide and 150 of the most frequent Class I HLA molecules in the human population, (ii) the predicted potential of the candidate peptide to be efficiently processed by the antigen processing machinery of the host infected cell, and (iii)

the predicted probability of the candidate peptide to be presented on the host infected cell surface, which, among other factors, takes binding and processing into account. For each protein a matrix of AP scores was generated. This matrix contains AP scores at each amino acid residue position. As a result, a matrix of aggregated AP scores for each amino acid residue-HLA pair is generated, which
5 can be represented as a heatmap for each protein.

To ascertain whether specific regions in a given viral protein were enriched with higher immunogenic scores as compared to the rest of the protein, a framework
10 was implemented based on a peak calling algorithm able to identify sub-regions of a protein which present higher immunogenicity than the rest of the protein. These regions were labelled as “hotspots”.

The entire map of AP score predictions of a given protein was used as input data. First, the AP scores for each amino acid within the protein sequence were summed across the 150 considered HLA alleles and the mean of these sums was
15 calculated. This mean represents the baseline or “waterline” for the peak detection algorithm. A local polynomial regression (LOESS) smoothing curve with a span of 10% was applied on the individual AP score sums (i.e., at amino acid level) thus generating a presentation potential “landscape” of the entire protein
20 sequence.

A rolling function was applied on the “landscape” to identify all local maxima within the protein. The default window size for the rolling function was set to nine amino acids (i.e., 9-mers), in agreement with the framework used to predict the presentation potential of each peptide within the protein. The identified peak
25 maxima represent the peak summits across the “landscape”. Only peak summits that are above the “waterline” (i.e., the baseline) were considered.

The peak summits were then input into a run length encoding (RLE) algorithm to identify the peak regions. A peak region was defined as the entire region flanking a peak summit on both sides that is consistently above the “waterline”.

The minimum required size of a peak region was set to 15 amino acids, to conjointly account for Class II HLAs, and the maximum allowed length was set to 100 amino acids, to meet the limitations of peptide synthesis. Peak regions that were shorter than the minimum required length were flanked on both sides with the remaining number of amino acids required to meet the minimum length. Then, the best window of 15 amino acids, with respect to the sum of AP scores across its constituting amino acids was kept as the corresponding hotspot. For peak regions exceeding 100 amino acids, but up to 150 amino acids in length, the same approach was used: the best window of 100 amino acids was kept. The peak regions longer than 150 amino acids were recursively split into shorter regions by re-calculating the baseline within the region (i.e., raising the “waterline”) and subsequent *de novo* detection of the peak summits and peak regions.

Finally, the amino acid sequences corresponding to the peak regions identified across the proteins are extracted, together forming the entire set of hotspots for a given viral species.

As explained above, the AP scores for a hotspot, once the hotspot is identified as containing likely immunogenic epitopes (AP score 0 to 1), can be combined and represented by the sum of their values. A given hotspot may contain multiple epitopes, and as such will have a summed AP score that reflects this. For example, a summed AP score of ~60 may indicate that a given hotspot contains many predicted epitopes, such that the hotspot is a desirable target for a potential vaccine. Furthermore, a hotspot that has a low summed AP score may still be desirable, as this may indicate the presence of a unique epitope sequence for a particular HLA type. This would be indicated by a summed AP score between 0 and 1, for a given hotspot. A low summed AP score that exceeds 1 may indicate the presence of a low number of likely immunogenic epitopes for a low number of HLA types. While hotspots such as these may not be widely immunogenic, their use to provoke immune responses in individuals with HLA types that would otherwise be difficult to target may be desirable. This analysis was conducted on the hotspot sequences identified and is represented below in Table 3. For

- example, SEQ ID No: 3 has an AP score sum of 62.66. This indicates the predicted presence of multiple epitopes. Whereas SEQ ID No: 9 has an AP score sum of 0.15. This indicates the predicted presence of fewer epitopes than in SEQ ID No: 3, but these epitopes may account for HLA types that would otherwise be unaccounted for. As such, a desirable vaccine may comprise hotspot sequences that have high AP scores (for example, sum of AP scores >10), as well as comprising hotspot sequences that have low AP scores (for example, sum of AP scores <1). A desirable vaccine may also comprise hotspot sequences with exclusively high or low AP scores.
- 10 Table 3: The sum of antigen presentation (AP) scores for the amino acid hotspot sequences identified to be broadly immunogenic across multiple Betacoronavirus subgenera, across multiple HLAs, or across unique HLAs.

SEQ ID No	Amino acid sequence	Sum of AP Scores
1	FVDGVPFVVSIGYHYKELGIVMNMDVDTHRYRLSLKDLLLLYAADP	15.18397404
3	NRATLQAIASEFSSLPSYAAYATAQEAY	55.48960739
5	QRVFNMYMPYVFTLLFQLCTFTKSTNSRIRASLPTTIKKN	62.65874973
7	TLLALREVKTIKVFTTVDNINLHTQVIDMSMTYGGQLG	1.169944824
9	LYRLVVDAYQAPPTI	0.148835658
11	GFSKFVTVQGPPGTGKSHFAIGLAIYYPTARVVY	5.636217613
13	LHVTYVPSQERNFTTAPAICHEGKAYFPREGVFVF	43.81698547
15	IMHVSFLVMFSGVPLWTVLYVVVLGRHVFWFTSLCTKRTVHVGDLVFHFSFQDA ALQTFMLDKEVFLRLKREISPETYARYLALFNKYKYSGPMDTA	1.131560943
17	TSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHNQDVNLHSSRLSFKELLVYAADPA MHAASGNLLLDKRTTCFSVAALTN	60.4826028
19	GPSDASALLYLEAKRLEALEAGKSKSAPKVITKKDAADAKKKMRHKRVATKAYNPT QAFGLRGP	0.949464351
22	GFSKFVTVQGPPGTGKSHFAIGLAIYYPTARVVY	5.636217613
24	TFSTFEEAALCTFLNKEMYLKLKRS	21.60787906
26	NVSVSRFNPSTWNRFRFGFTEQSVFKPQP	0.883974105
28	IVNVSLVKPTVYVYSRVKN	45.45080774
30	ATIPLQASLPFGWLVIQVAFQVQSAVKIALLNKRWQLALYKGFQFI	47.19962358
31	GETALALLLDRLNQL	50.04202975
33	VPIGGTTVVRPLVED	0.801315543
35	KINQALHGANLRQDESVSLSY	1.483679342

37	FAYSNRNRFLYIIKLVFL	51.09517394
39	LHVTYVPSQERNFTTAPAICHEGKAYFPREGVFVF	43.81698547
41	LSPRWYFYFLGTGPEA	36.40144149
43	QRVFNNYMPYVFTLLFQLCTFTKSTNSRIRASLPTTIKLN	62.65874973
45	ADLYKLMGHFSWWTAFVT	28.1841187
46	FAGTHITMTTVYHITSQIQLSLLKVT	45.34786326
48	RVAGSITMRSGSSPAL	0.30003004
49	AAIHYLNSKIVDLAQHFS	15.4464556
51	VYMPASWVMRIMTWLEL	76.81897834
53	ECTYISRFISKISLVSV	1.68622931
55	LSVIVQPTHYVRVTFSDPNMWYLR	1.886049223
57	LPKEITVATSRTLSTYKLG	48.02001639
59	ASSLFAQFASVTNVTTAMGKALGR	3.077867951
61	PNNTASWFTALTQHGKEELRFPR	3.171322912
63	QFPILAEAPTAGAFFGSRLELAKVQNLSG	9.83728293
64	QLGNSFGAISSSINEIFSRLD	1.532888563
67	FVDGVPFVVSIGYHYKELGIVMNMVDVTHRYRLSLKDLLLYAADP	15.18397404
69	HYYSDFWIKIGSRKSARL	4.959632058
71	NLDVISSIVRQLFK	35.08483405
73	NLDMISSIVRQLFK	2.046477734
75	SELVIGAVIIRGHLRMAG	27.49269719
77	KIQTAPHISLAMLDIQPE	4.61139961
79	GPMLRLIHLVSLWTF	1.743091184
81	VVPYYSWFSGITQFQKGKEFEFAEGQGVPPIAPGVPAE	11.07794289
83	KSAGFPFNKFGKARLYYESLSY	1.619626118
85	ALTTFMITKDSYCKLNLSLSDVAFNRYL	23.45692935
87	RQKRTATKQYNVTQAFGRR	32.55333944
89	NRATLQAIASEFSSLPSYAAAYATAQEAY	55.48960739
91	GVAFLAVFQSATKIIALNKRWQLAL	6.400134665
93	QNKGLFYTVVYKNMAVYRSLTF	2.77397747
98	LFSTVFPLTSFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVNLHSSRLSFKELL VYAADPAMHAASGNLLDKRTTCFSAALTNNVAFQTVKPGNF	60.4826028
106	RHINAQVAKSHNIALIWNVKDFMSLSEQLRKQIRSAAKNNLPFKLTCAATTRQVVNV VTTKIALKGGKIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVM	3.506818961
108	EHFIETISLAGSYKDWSSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVITFDNL KTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGGQFGPTY	1.069780381
110	AVVYRGTTTYKLVNGDYFVLTSHTVMPLSAPTLVPQEHYVRITGLYPTLNISDEFSS NVANYQKVGMMQKYSTLQGGPGTGKSHFAIGLALYYPSARIVYT	44.65342515
112	LPKEITVATSRTLSTYKLG	48.02001639

113	AIGKIQDLSSTASALGKLQDVVNHNAQALNTLVKQLSSKFGAISSVLNDILSRLDKV EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKM	59.89176692
114	EVVGNVILKPSDEGVKVTQELGHEDLMAAYVENTSITIKKPNELSLALGLKTIATHG	2.651375671
116	DRYFEIYDGGCLNASEVIVNNLDKSAGHPFNKFGKARVYYESLSYQEDELFAATK RNVLPITITQMNLYAISAKNRARTVAGVSILSTMTNRQYHQKMLKSMA	3.821131468
118	YNSDYASKGYIANISAFQQVLYYQNNVFMSEAKCWENDINHGPFHFCQSHTMLV KMDGDDVYLPYPDPSRILGAGCFVDDLLKTDSVLLIERFVSLAIDAY	1.354392237
120	EYKGPVTDVFKYKTSYTTTIKPVSYKLDGVTYTEIEPKLDGYYKKNAYVENTSITIK KPNELSLALGLKTIATHGIAAINSVPWSKILAYVKPFLGQAA	2.708784475
122	AEHFHETVSLAGSYKDWSSYSGQRTLGVEFLKRGDKIVYHTVGNVIEFHMEGEVL PLEKLTLLALREVKTIKVFTTVDNINLHTQVIDMSMTYGQQLG	1.169944824
124	NARDGCVPLNIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQVVDADSK IVQLSEINMDNSQNLAWPLIVTALRANSVAVKLQNNELSPVALR	0.94582376
126	PIEFHLDGEVPLDKLSLLSLREVKTIKVFTTVDNTNLHTQLVDMSVTYGQHFPGTY	1.321187163
128	KMADQAMTQMYKQARSEDKRAKVTSAMQTMFLTMLRKLDNDALNNIINNARDGCV PLNIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQVQV	1.221276774
130	VSLIWNVKDYMSLSEQLRKQIRSAKKNNIPFRLTCATTRQVV	3.506818961
132	LAYMLFTKFFYLLGLSAIMQVFFGYFASHFISN	2.245702591
134	VLKLLKSLNVAKSEFDRDAAMQRKLEKMAQAMTQMYKQARSEDKRAKVTSAM QTMFLTML	60.22132275
135	KMADQAMTQMYKQARSEDKRAKVTSAMQTMFLTMLRKLDNDALNNIINNARDGCV PLNIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQVQV	1.221276774
137	HHANEYRQYLDAYNMMISAGFSLWIYKQFDYTNLWNTFTRLQSLNVAYNVNKG HFDGH	7.589886009
139	LHVTYVPSQERNFTTAPAICHEGKAYFPREGVVF	43.81698547
141	TSFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVNLHSSRSLFKELLVYAADPA MHAASGNLLDKRTTCFSVAALTN	60.4826028
143	VFISPYNSQNAVASKILGLPTQTVDSSQGSEYDYVFTQTTETAHSCNVNRFNVAITR AKIGILCIMSDDRLYDKLQFTSLEIPRRNVATLQAENVTLGFKD	41.97833518
145	EYKGPVTDVFKYKTSYTTTIKPVSYKLDGVTYTEIEPKLDGYYKKNAYVENTSITIK KPNELSLALGLKTIATHGIAAINSVPWSKILAYVKPFLGQAA	2.708784475
147	NARDGCVPLNIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQVVDADSK IVQLSEINMDNSQNLAWPLIVTALRANSVAVKLQNNELSPVALR	0.94582376
149	AEHFHETVSLAGSYKDWSSYSGQRTLGVEFLKRGDKIVYHTVGNVIEFHMEGEVL PLEKLTLLALREVKTIKVFTTVDNINLHTQVIDMSMTYGQQLG	1.169944824
150	YNSDYASKGYIANISAFQQVLYYQNNVFMSEAKCWENDINHGPFHFCQSHTMLV KMDGDDVYLPYPDPSRILGAGCFVDDLLKTDSVLLIERFVSLAIDAY	1.354392237
152	DRYFEIYDGGCLNASEVIVNNLDKSAGHPFNKFGKARVYYESLSYQEDELFAATK RNVLPITITQMNLYAISAKNRARTVAGVSILSTMTNRQYHQKMLKSMA	3.821131468
153	PIEFHLDGEVPLDKLSLLSLREVKTIKVFTTVDNTNLHTQLVDMSVTYGQHFPGTY	1.321187163
155	AEHFHETVSLAGSYKDWSSYSGQRTLGVEFLKRGDKIVYHTVGNVIEFHMEGEVL PLEKLTLLALREVKTIKVFTTVDNINLHTQVIDMSMTYGQQLG	1.169944824
157	HHANEYRQYLDAYNMMISAGFSLWIYKQFDYTNLWNTFTRLQSLNVAYNVNKG HFDGH	7.589886009
159	LAYMLFTKFFYLLGLSAIMQVFFGYFASHFISN	2.245702591
161	TKRNVLPITITQMNLYAISAKNRARTVAGVSI	2.529654529
163	YNSDYASKGYIANISAFQQVLYYQNNVFMSEAKCWENDINHGPFHFCQSHTMLV KMDGDDVYLPYPDPSRILGAGCFVDDLLKTDSVLLIERFVSLAIDAY	1.354392237
165	GLDSLDSYPALETIQVTISSYKDLT	1.579382094

167	ALTTFMITKDSYCKLKNLSLSDVAFNRYL	23.45692935
169	PAYSFLPGVYSVFYLYLTFYFTNDVSLAHLQWFAMFSPIVPFWIT	7.380232066
171	NARDGCVPLNIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQVVDADSK IVQLSEINMDNSQNLAWPLIVTALRANSVVKLQNNELSPVALR	0.94582376
172	EVVGNVILKPSDEGVKVTQELGHEDLMAAYVENTSITIKKPNELSLALGLKTIATHG	2.651375671
174	TSFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVNLHSSRSLFKELLVYAADPA MHAASGNLLDKRTTCFSVAALTN	60.4826028
176	EKGPFHFCSQHTMLVKMDGDEVLYPYDPSPRILGAGCFVDDLLKTDSVLLIERFVSL AIDAYPLVHHENPEYQNVFRVYLEYIKKLYNDLGNQILDSYSV	1.716533355
178	KMADQAMTQMYKQARSEDKRAKVTSAMQTMFLTMLRKLNDNALNNIINNARDGCV PLNIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQVV	1.221276774
180	DRYFEIYDGGCLNASEVIVNNLDKSAGHPFNKFGKARVYYESLSYQEDELFAFMTK RNVLPITQMNLKYAISAKNRARTVAGVSILSTMNTRQYHQKMLKSMA	3.821131468
182	RTVAGVSICSTMTNRQFHQKLLKSIAATRGAATVIGTSKFYGGWHNMLKTVYSDVE	7.987563834
184	EYKGPVTDVFKYKTSYTTTIKPVSYKLDGVTYTEIEPKLDGYKKNAYVENTSITIK KPNELSLALGLKTIATHGIAAINSVPWSKILAYVKPFLGQAA	2.708784475
186	VSLIWNVKDYMSLSEQLRKQIRSAKKNNIPFRLTCATTRQVV	3.506818961
188	TFSTFEEAALCTFLNKEMYLKLRE	21.60787906
190	VLKKLKSLNVAKSEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAM QTMFLTML	60.22132275
193	AVVYRGTTTYKLVNGDYFVLTSHTVMPLSAPTLPQEHYVRITGLYPTLNISDEFSS NVANYQKVGMMQKYSTLQGGPGTGKSHFAIGLALYPSARIVYT	44.65342515
197	EHFIETISLAGSYKDWYSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVITFDNL KTLISLREVRTIKVFTTVDNINLHTQVVDMSMTYQQQFGPTY	1.069780381
199	LFSTVFPLTSFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVNLHSSRSLFKELL VYAADPAMHAASGNLLDKRTTCFSVAALTNNAFQTVKPGNF	60.4826028
203	AIGKIQDSLSSASALGKLQDVVNHNAQALNTLVKQLSSKFGAISSVLNDILSRDKV EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKM	59.89176692
206	LPKEITVATSRTLSYYKLG	48.02001639
208	RHINAQVAKSHNIALIWNVKDFMSLSEQLRKQIRSAKKNNLPFKLTATTRQVVNV VTTKIALKGGKIVNNWLKLIKVTLVFLVAAIFYLITPVHVM	3.506818961
209	DVLISSIVRQLFKPL	1.614550874
211	LHVTYVPSQERNFTTAPAICHEGKAYFPREGVVF	43.81698547
213	AEHFIEIVSLAGSYKDWYSYSGQRTLGVEFLKRGDKIVYHTVGNVIEFHMEGEVL PLEKLTLLALREVKTIVFTTVDNINLHTQVIDMSMTYGGQLG	1.169944824
215	LPKEITVATSRTLSYYKLG	48.02001639
217	DRYFEIYDGGCLNASEVIVNNLDKSAGHPFNKFGKARVYYESLSYQEDELFAFMTK RNVLPITQMNLKYAISAKNRARTVAGVSILSTMNTRQYHQKMLKSMA	3.821131468
218	LAYMLFTKFFYLLGLSAIMQVFFGYFASHFISN	2.245702591
219	SVLQATLSEFSLASAELEAAQSSYQKALNSGDASPOVLKALQKAVNIAKNAYEK DKAVARKLERMAEQAMTSMYKQARAEDKKAKIVSAMQTM	4.779086126
221	KMADQAMTQMYKQARSEDKRAKVTSAMQTMFLTMLRKLNDNALNNIINNARDGCV PLNIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQVV	1.221276774
223	HHANEYRQYLDAYNMMISAGFSLWYKQFDYTNLWNTFTRLQSLENVAYNVVNKG HFDGH	7.589886009
225	NRATLQAIASEFSSLPSYAAYATAQEAY	55.48960739
227	YNSDYASKGYIANISAFQQVLYYQNNVFMSEAKCWENDINHGPHEFCSQHTMLV KMDGDDVLYPYDPSPRILGAGCFVDDLLKTDSVLLIERFVSLAIDAY	1.354392237

229	TSLEIPRRNVATLQAENVNGLFKDCSKIITGLHPTQAPTHLSVDIKFKTEGLCVDIPGI PKDMTYRRLISMMGMNYQVNGYPNMFITREEAIRHVRAW	0.98370983
231	GSNLEEEILLMLASLFGTYTWTTVLSMA	1.738317558
233	LSYFVASFRLFARTRS	2.255780642
237	FVDGVPFVVSIGYHYKELGIVMNMVDVTHRYRLSLKDLLLYAADP	15.18397404
239	EYKGPVTDVIFYKETSYYTTTIKPVSYKLDGVITYEIEPKLDGYKKDNAYVENTSITIK KPNELSLALGLKTIATHGIAAINSPWSKILAYVKPFLGQAA	2.708784475
241	SSYMAISLKFTTSLSLHKLQTLVLK	0.898281036
243	RTVAGVSICSTMTNRQFHQKLLKSIAATRGATVIGTSKFYGGWHNMLKTVYSDVE	7.987563834
245	VFISPYNSQNAVASKILGLPTQTVDSQSGSEYDYVIFTQTTETAHSCNVNRFNVAITR AKIGILCIMSDDRDLQFTSLEIPRRNVATLQAENVNGLFKD	41.97833518

Example 3

To ascertain whether there were highly immunogenic regions that were shared by multiple Betacoronavirus subgenera, a framework was developed to identify and cluster hotspots that were common to more than one viral species.

For that, all possible pairwise combinations of identified hotspots were generated across all proteins and all Betacoronaviruses.

$$\{ \{ x, y \} \mid x, y \in O \wedge x \neq y \}$$

with x and y being two non-identical hotspots and O the entire set of Betacoronaviruses. Next, the hotspot sequences were converted into ordered vectors of amino acid frequencies. For instance, the frequency vector of hotspot x would be denoted as:

$$F_x = \{ (fa_1, \dots, fa_n) \mid a_i \in A \wedge 1 \leq i \leq |A| \}$$

where A is the complete, ordered set of amino acids compiled based on the entire protein sequence input data and fa_i is the frequency of amino acid i within the hotspot sequence. The similarity between the two amino acid frequency vectors within each pair was assessed by means of Pearson correlation coefficient (ρ):

$$\rho_{(x,y)} = \text{corr}(Fx, Fy)$$

This allowed for the comparison of the amino acid proportions between two hotspot sequences. Only hotspot pairs with a $\rho > 0.95$ were kept for further analysis:

$$C = \{ (x,y) \mid \rho_{(x,y)} > 0.95 \}$$

The chosen cutoff value is a stringent threshold in itself. However, it still allowed for point mutations and/or differences in the flanking regions (i.e., shifts in the start and/or end coordinates of the hotspot sequence) of the two hotspot sequences being compared.

- 10 Potential false positives can arise as (1) hotspot sequences with the same amino acid composition, but in a different order (e.g., mirrored or shuffled) or (2) very short hotspot sequences (e.g., 15 amino acids) for which the amino acid frequency profile is quite flat (i.e., a frequency vector with most values being 0 and/or 1). To discard false positives, the pairwise hotspot sequence similarity was calculated
- 15 for the entire set of pairs with a $\rho > 0.95$:

$$S_{(x,y)} = \{ 1 - \text{dist}(x, y) / \max(\text{dist}(x, y)) \mid (x, y) \in C \}$$

- The distance *dist* between sequence x and y is calculated as the restricted Damerau-Levenshtein distance. The sequence similarity is a value between 0 and 1, with 0 denoting two completely different character sequences and 1 denoting
- 20 identical character sequences.

Next, the difference between the Pearson correlation coefficient and sequence similarity was calculated for all considered hotspot pairs:

$$D_{(x,y)} = \{ \rho_{(x,y)} - S_{(x,y)} \mid (x, y) \in C \}$$

- Based on the distribution of $D_{(x,y)}$, the 99th percentile was used as a cutoff to
- 25 discard hotspot pairs presenting extreme differences (i.e., equal to or above the

99th percentile). In other words, pairs of hotspots with a high ρ but low $S_{(x,y)}$ were considered false positives and discarded. As such, the final set of hotspots was defined as:

$$H = \{ (x, y) \mid D_{(x,y)} < P_{99}(D) \wedge (x, y) \in C \}$$

- 5 These hotspots were then placed under a directed graph representation, where each hotspot was represented by a vertex and an edge between two vertices was defined for each $(x, y) \in H$. Finally, clusters of homologous hotspots were identified by extracting all connected vertices within the graph. Due to the nature of this implementation, the minimum size of a hotspot cluster was two. In other words, a
- 10 hotspot was considered for further analysis if it is common to at least two Betacoronavirus subgenera.

Example 4

- The clusters of homologous hotspots were converted to the FASTA format and the hotspot sequences were aligned using Clustal Omega, a third party tool for
- 15 multiple sequence alignment (MSA), using default parameters.

- The output of the MSA tool was then used to extend individual hotspot sequences (i.e., per viral species) where possible, such that the “real estate” of the homologous hotspots is maximized. This was done by replacing the flanking dashes (“-”) output by Clustal Omega (i.e., placeholders used during the alignment
- 20 of the amino acid sequences) surrounding a hotspot sequence, with the actual amino acids at those positions taken from the original, corresponding protein sequence.

- The rationale behind this step was that minor differences in the baseline used in peak detection (i.e., “waterline”) would cause a hotspot region within a given
- 25 protein and a given viral species to be trimmed left and/or right with respect to the same hotspot region in an analogous protein in a different viral species. However,

such small differences in the “waterline” level should not have a notable impact on the overall hotspot region, with respect to its presentation potential.

The clusters of extended hotspots were then re-aligned using the same MSA third party tool, using the default parameter settings. Based on the sequence of
5 consensus symbols that Clustal Omega outputs, each cluster of homologous hotspots was processed to identify which hotspot sub-regions are the most conserved across the viral species accounted for within the cluster. An amino acid position within a hotspot was considered conserved if designated with an “*” or a “:” in the Clustal Omega consensus symbol sequence.

10 Next, the consensus symbol sequence was transformed into a binary vector, with all conserved amino acid positions converted to a 1 and all other amino acid positions converted to a 0. Run length encoding (RLE) was then used to identify all the conserved regions within a cluster of homologous hotspots. In other words, the regions of consecutive 1s were delineated. Only conserved regions that were
15 at least 15 amino acids in length were kept. Note that one hotspot cluster can yield one, several, or no highly conserved regions.

The level of conservation for each highly conserved hotspot region was identified, and is represented in Figure 3. More immunogenic sequences across a higher number of Betacoronavirus subgenera were considered desirable vaccine
20 candidates.

Example 5

The sets of protein sequences for each of the five Betacoronavirus subgenera were assessed for their potential to be presented on the host infected cell surface. This allowed for each amino acid within a protein sequence to be assigned an
25 antigen presentation (AP) score. Based on these AP scores, regions within the viral proteins that show high predicted presentation potential were identified. These regions were labelled as “hotspots”.

A total of 16,000 hotspots across all five Betacoronavirus subgenera were identified, with more than 4,000 unique hotspots identified across all species. The minimum length of a hotspot region was set to 15 amino acids, to account for Class II HLA alleles and the maximum length was set to 100 to meet the limitations of peptide manufacturing. The number of hotspots per protein ranged from 1 to 27, depending on protein length and distribution of AP scores within the protein sequence. There were no notable outliers present in the data, with respect to the number of hotspots and their length distribution.

The entire set of hotspots was used to check if there were any homologous protein regions, with a high presentation potential across multiple Betacoronavirus subgenera, which could lead to the identification of cross-reactive epitopes. Clusters of highly similar hotspots across different species were therefore identified.

Out of those, 132 hotspot clusters indicated broad cross-reactivity across the five Betacoronavirus subgenera for different HLA types. These hotspot clusters can be found in the sequences of claim 1.

The highly conserved hotspot regions within each hotspot cluster were then identified. Taken together, these results show that there is a substantial set of protein subsequences, homologous to multiple Betacoronavirus subgenera that show high sequence conservation, thus representing potential candidates for a broadly protective T-cell vaccine against all Betacoronavirus subgenera.

Example 6

To assess the immunogenicity of the hotspots and availability of CD4+ and CD8+ T cells that could respond to these predictions, i) peptide pools (Table 4) comprising the identified peptides and ii) single peptides that cover the sequences with best AP scores and HLA-binding properties were tested.

To reflect the post pandemic situation and utility of the vaccines for the population, donors were selected who received ≥ 2 doses of mRNA vaccines (Moderna or Pfizer/BioNTech) and had recovered from COVID-19.

The protocol for T-cell line stimulation (ex vivo vaccination) starts with collecting
5 Peripheral Blood Mononuclear Cells (PBMCs), divide them in two equal parts of
approximately 5-10 million cells each, and stimulate one part (1h) with
immunoreactive Betacoronavirus peptides (used at a final concentration of 2.25
 μg peptide/ml). Stimulated PBMCs are mixed with the unstimulated portion and
rested in culture for three days and then expanded in presence of IL-2 and IL-7. T
10 cells were restimulated two more times using irradiated peptide loaded autologous
PBMCs, and then analysed by flow cytometry. Following ex vivo vaccination with
SARS-CoV-2 Spike Peptivator peptide pool or partially agonistic control B2 pool,
>65% of expanded T cells (both CD8+ and CD4+ T cells) were activated when re-
challenged with the specific stimulus. Shown in Figure 6 are activated T cells
15 measured by the combined expression of CD137 and IFNG markers.

Pools were tested on T cell lines that had been expanded as shown in Figure 6.
Single peptide CD4+ and CD8+ T cell responsiveness were tested on PBMC
directly from donors. Results showed the existence of T cell responsiveness
towards the peptide pools (Table 5- as well as single peptides. The generation of
20 T cell lines represented the expansion of CD4+ and CD8+ T cells following in
response to peptides present in the vaccine constructs, and served as an in vitro
vaccine response assay. The later re-challenge allowed quantification of vaccine
peptide responsive CD4+ and CD8+ T cells and analysis of T cell function. Robust
T cell functionality is most often signified by multiple response patterns, called
25 polyfunctionality. The polyfunctionality is shown in Figures 7-10.

The peptides of Table 4 and Table 5 are all found within the sequences of claim 1,
and show that the claimed hotspot sequences have immunogenic epitope
sequences within them. Vaccines constructed from these hotspot sequences,
which in turn comprise the immunogenic peptides of Table 4 and Table 5, can

therefore stimulate a broad and effective immune response across a diverse spectrum of human HLA alleles and across multiple Betacoronavirus subgenera.

Table 4: The minimal epitope Betacoronavirus peptides used for generation of T cell lines (*ex vivo* vaccination) and testing of peptide pools derived from hotspots.

5 The corresponding Sequence ID Number (SEQ ID No.) of the hotspot within which the minimal epitope peptide appears and Construct ID are provided for all the peptides.

Peptide sequence	Sequence length	Protein	Seq ID No. peptide is found within	Construct(s) ID	Top MHC-I binding prediction (NIP)
FVDGVPFVV	9	ORF1ab	1, 67, 237	t_cell_max_marge-construct13, t_cell_d4-mc20-393, t_cell_d2-mc1000- 1717	HLA-A-02-01
AIASEFSSL	9	ORF1a	3, 89, 225	t_cell_max_marge-construct13, t_cell_d4-mc20-393, t_cell_d2-mc1000- 1717	HLA-B-15-02
FSSLPSYAA	9	ORF1a	3, 89, 225	t_cell_max_marge-construct13, t_cell_d4-mc20-393, t_cell_d2-mc1000- 1717	HLA-B-15-02

SEFSSLPSY	9	ORF1a	3, 89, 225	t_cell_max_marge- construct13, t_cell_d4-mc20- 393, t_cell_d2- mc1000- 1717	HLA-B-15- 01; HLA- B-15-02; HLA-B-44- 02
TLQAIASEF	9	ORF1a	3, 89, 225	t_cell_max_marge- construct13, t_cell_d4-mc20- 393, t_cell_d2- mc1000- 1717	HLA-B-15- 01; HLA- B-15-02
IFVDGVPFV	9	ORF1ab	17, 141, 174	t_cell_max_marge- construct13, t_cell_d5-mc10- 977, t_cell_d5- mm20- 620-wf	HLA-A-02- 01
KIFVDGVPF	9	ORF1ab	17, 141, 174	t_cell_max_marge- construct13, t_cell_d5-mc10- 977, t_cell_d5- mm20- 620-wf	HLA-B-15- 01; HLA- B-15-02
LLDKRTTCF	9	ORF1ab	17, 141, 174	t_cell_max_marge- construct13, t_cell_d5-mc10- 977, t_cell_d5- mm20- 620-wf	HLA-B-15- 02

LVYAADPAM	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-B-15-01; HLA-B-15-02
ASGNLLLDK	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-11-01
LSFKELLY	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-01-01; HLA-A-11-01; HLA-A-29-02; HLA-B-15-01; HLA-B-15-02
RLSFKELLV	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-02-01
TSFGPLVRK	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-03-01; HLA-A-11-01

VVSTGYHFR	9	ORF1ab	17, 141, 174	t_cell_max_marge- construct13, t_cell_d5-mc10- 977, t_cell_d5- mm20- 620-wf	HLA-A-03- 01; HLA- A-11-01
FVVSTGYHF	9	ORF1ab	17, 141, 174	t_cell_max_marge- construct13, t_cell_d5-mc10- 977, t_cell_d5- mm20- 620-wf	HLA-A-23- 01; HLA- A-29-02; HLA-B-15- 01; HLA- B-15-02
NLHSSRLSF	9	ORF1ab	17, 141, 174	t_cell_max_marge- construct13, t_cell_d5-mc10- 977, t_cell_d5- mm20- 620-wf	HLA-B-08- 01; HLA- B-15-01; HLA-B-15- 02
SSRLSFKEL	9	ORF1ab	17, 141, 174	t_cell_max_marge- construct13, t_cell_d5-mc10- 977, t_cell_d5- mm20- 620-wf	HLA-B-08- 01
VPFVVSTGY	9	ORF1ab	17, 141, 174	t_cell_max_marge- construct13, t_cell_d5-mc10- 977, t_cell_d5- mm20- 620-wf	HLA-B-15- 02

FVDGVPFVV	9	ORF1ab	1, 67, 237	t_cell_max_marge-construct13, t_cell_d4-mc20-393, t_cell_d2-mc1000- 1717	HLA-A-02-01
LLLDRLNQL	9	N	31	t_cell_d2-mc1000-1717	HLA-A-02-01; HLA-B-08-01
VATSRTLSTY	9	M	57, 112, 206, 215	t_cell_d2-mc1000-1717, t_cell_d4-mc20-393, t_cell_d3-mc10-1375, t_cell_d3-mm11- 104	HLA-A-29-02; HLA-B-15-01; HLA-B-15-02
ATSRTLSTYY	9	M	57, 112, 206, 215	t_cell_d2-mc1000-1717, t_cell_d4-mc20-393, t_cell_d3-mc10-1375, t_cell_d3-mm11- 104	HLA-A-01-01; HLA-A-03-01; HLA-A-11-01; HLA-A-29-02; HLA-B-15-01
LPKEITVAT	9	M	57, 112, 206, 215	t_cell_d2-mc1000-1717, t_cell_d4-mc20-393, t_cell_d3-mc10-1375, t_cell_d3-mm11- 104	HLA-B-07-02

TSRTLSYYK	9	M	57, 112, 206, 215	t_cell_d2-mc1000-1717, t_cell_d4-mc20-393, t_cell_d3-mc10-1375, t_cell_d3-mm11- 104	HLA-A-03-01; HLA-A-11-01
LSPRWYFY	9	N	41	t_cell_d2-mc1000-1717	HLA-A-01-01; HLA-A-29-02
SPRWYFY	9	N	41	t_cell_d2-mc1000-1717	HLA-B-07-02; HLA-B-08-01
YYLGTGPEA	9	N	41	t_cell_d2-mc1000-1717	HLA-A-23-01; HLA-A-24-02; HLA-A-29-02
AIASEFSSL	9	ORF1a	3, 89, 225	t_cell_max_marge-construct13, t_cell_d4-mc20-393, t_cell_d2-mc1000- 1717	HLA-B-15-02
FSSLPSYAA	9	ORF1a	3, 89, 225	t_cell_max_marge-construct13, t_cell_d4-mc20-393, t_cell_d2-mc1000- 1717	HLA-B-15-02

SEFSSLPSY	9	ORF1a	3, 89, 225	t_cell_max_marge-construct13, t_cell_d4-mc20-393, t_cell_d2-mc1000- 1717	HLA-B-15-01; HLA-B-15-02; HLA-B-44-02
TLQAIASEF	9	ORF1a	3, 89, 225	t_cell_max_marge-construct13, t_cell_d4-mc20-393, t_cell_d2-mc1000- 1717	HLA-B-15-01; HLA-B-15-02
FTALTQHGK	9	N	61	t_cell_d2-mc1000-1717	HLA-A-11-01
SELVIGAVI	9	M	75	t_cell_d2-mc1000-1717	HLA-B-40-01; HLA-B-44-02
ALCTFLLNK	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-A-03-01; HLA-A-11-01
FLLNKEMYL	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-A-02-01; HLA-B-08-01
LLNKEMYLK	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-A-03-01; HLA-A-11-01

EEAALCTFL	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-B-40-01
FEEAALCTF	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-B-40-01
LNKEMYLKL	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-B-08-01
TFLLNKEMY	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-A-29-02
MPASWVMRI	9	ORF1a	51	t_cell_d2-mc1000-1717	HLA-B-07-02; HLA-B-15-02
SWVMRIMTW	9	ORF1a	51	t_cell_d2-mc1000-1717	HLA-A-23-01; HLA-A-24-02
VYMPASWVM	9	ORF1a	51	t_cell_d2-mc1000-1717	HLA-A-23-01; HLA-A-24-02; HLA-B-15-02
WVMRIMTWL	9	ORF1a	51	t_cell_d2-mc1000-1717	HLA-A-02-01

VATSRTLSTY	9	M	57, 112, 206, 215	t_cell_d2-mc1000-1717, t_cell_d4-mc20-393, t_cell_d3-mc10-1375, t_cell_d3-mm11- 104	HLA-A-29-02; HLA-B-15-01; HLA-B-15-02
ATSRTLSTYY	9	M	57, 112, 206, 215	t_cell_d2-mc1000-1717, t_cell_d4-mc20-393, t_cell_d3-mc10-1375, t_cell_d3-mm11- 104	HLA-A-01-01; HLA-A-03-01; HLA-A-11-01; HLA-A-29-02; HLA-B-15-01
LPKEITVAT	9	M	57, 112, 206, 215	t_cell_d2-mc1000-1717, t_cell_d4-mc20-393, t_cell_d3-mc10-1375, t_cell_d3-mm11- 104	HLA-B-07-02
TSRTLSTYYK	9	M	57, 112, 206, 215	t_cell_d2-mc1000-1717, t_cell_d4-mc20-393, t_cell_d3-mc10-1375, t_cell_d3-mm11- 104	HLA-A-03-01; HLA-A-11-01
AMQTMLFTM	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977,	HLA-A-02-01; HLA-B-15-01;

				t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-02
MQTMLFTML	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-08-01; HLA-B-15-01; HLA-B-15-02
KRAKVTSAM	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-02
KVTSAMQTM	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-02
MADQAMTQM	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-02

TSAMQTMLF	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-01; HLA-B-15-02
AAMQRKLEK	9	ORF1a	134, 190	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-A-11-01
LEKMADQAM	9	ORF1a	134, 190	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-02
SEFDRDAAM	9	ORF1a	134, 190	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-02; HLA-B-40-01; HLA-B-44-02
SLNVAKSEF	9	ORF1a	134, 190	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-01; HLA-B-15-02
VLKKLKKSL	9	ORF1a	134, 190	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-08-01
KQIRSAKK	9	ORF1ab	130, 186	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-A-03-01; HLA-A-11-01

MSLSEQLRK	9	ORF1ab	130, 186	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-A-03-01; HLA-A-11-01
QLRKQIRSA	9	ORF1ab	130, 186	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-08-01
AMQTMLFTM	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-A-02-01; HLA-B-15-01; HLA-B-15-02
MQTMLFTML	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-08-01; HLA-B-15-01; HLA-B-15-02
KRAKVTSAM	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-02
KVTSAMQTM	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-	HLA-B-15-02

				620-wf, t_cell_d4-mm11- 193-wf	
MADQAMTQM	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-02
TSAMQTMLF	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-01; HLA-B-15-02
FVDGVPFVV	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-02-01
IFVDGVPFV	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-02-01
KIFVDGVPF	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-	HLA-B-15-01; HLA-B-15-02

				977, t_cell_d5-mm20- 620-wf	
LLDKRTTCF	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-B-15-02
LVYAADPAM	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-B-15-01; HLA-B-15-02
ASGNLLLDK	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-11-01
LSFKELLY	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-01-01; HLA-A-11-01; HLA-A-29-02; HLA-B-15-01; HLA-B-15-02
RLSFKELLV	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-	HLA-A-02-01

				977, t_cell_d5-mm20- 620-wf	
TSFGPLVRK	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-03-01; HLA-A-11-01
VVSTGYHFR	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-03-01; HLA-A-11-01
FVSTGYHF	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-23-01; HLA-A-29-02; HLA-B-15-01; HLA-B-15-02
NLHSSRLSF	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-B-08-01; HLA-B-15-01; HLA-B-15-02
SSRLSFKEL	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-	HLA-B-08-01

				977, t_cell_d5-mm20- 620-wf	
VPFVVSTGY	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-B-15-02
AMQTMLFTM	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-A-02-01; HLA-B-15-01; HLA-B-15-02
MQTMLFTML	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-08-01; HLA-B-15-01; HLA-B-15-02
KRAKVTSAM	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-02
KVTSAMQTM	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977,	HLA-B-15-02

				t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	
MADQAMTQM	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-02
TSAMQTMLF	9	ORF1a	128, 135, 178, 221	t_cell_d4-mc20-393, t_cell_d5-mc10-977, t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-01; HLA-B-15-02
ALCTFLLNK	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-A-03-01; HLA-A-11-01
FLLNKEMYL	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-A-02-01; HLA-B-08-01
LLNKEMYLK	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-A-03-01; HLA-A-11-01

EEAALCTFL	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-B-40-01
FEEAALCTF	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-B-40-01
LNKEMYLKL	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-B-08-01
TFLLNKEMY	9	ORF1a	24, 188	t_cell_d2-mc1000-1717, t_cell_d5-mm20- 620-wf	HLA-A-29-02
FVDGVPFVV	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-02-01
IFVDGVPFV	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-02-01
KIFVDGVPF	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-	HLA-B-15-01; HLA-B-15-02

				977, t_cell_d5-mm20- 620-wf	
LLDKRTTCF	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-B-15-02
LVYAADPAM	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-B-15-01; HLA-B-15-02
ASGNLLLDK	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-11-01
LSFKELLY	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-01-01; HLA-A-11-01; HLA-A-29-02; HLA-B-15-01; HLA-B-15-02
RLSFKELLV	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-	HLA-A-02-01

				977, t_cell_d5-mm20- 620-wf	
TSFGPLVRK	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-03-01; HLA-A-11-01
VVSTGYHFR	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-03-01; HLA-A-11-01
FVSTGYHF	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-A-23-01; HLA-A-29-02; HLA-B-15-01; HLA-B-15-02
NLHSSRLSF	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-B-08-01; HLA-B-15-01; HLA-B-15-02
SSRLSFKEL	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-	HLA-B-08-01

				977, t_cell_d5-mm20- 620-wf	
VPFVVSTGY	9	ORF1ab	17, 141, 174	t_cell_max_marge-construct13, t_cell_d5-mc10-977, t_cell_d5-mm20- 620-wf	HLA-B-15-02
AAMQRKLEK	9	ORF1a	134, 190	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-A-11-01
LEKMADQAM	9	ORF1a	134, 190	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-02
SEFDRDAAM	9	ORF1a	134, 190	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-02; HLA-B-40-01; HLA-B-44-02
SLNVAKSEF	9	ORF1a	134, 190	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-15-01; HLA-B-15-02
VLKKLKKSL	9	ORF1a	134, 190	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-08-01

KQIRSAKK	9	ORF1ab	130, 186	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-A-03-01; HLA-A-11-01
MSLSEQLRK	9	ORF1ab	130, 186	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-A-03-01; HLA-A-11-01
QLRKQIRSA	9	ORF1ab	130, 186	t_cell_d5-mm20-620-wf, t_cell_d4-mm11- 193-wf	HLA-B-08-01

5

10 Table 5: The minimal epitope Betacoronavirus peptides used to test CD4+ and CD8+ T cell immune responses towards single peptides directly *ex vivo* from healthy donors. The corresponding Sequence ID Number (Seq ID No.) of the hotspot within which the minimal epitope peptide appears and Construct ID are provided for all the peptides.

Peptide Sequence	Sequence Length	Seq ID No. peptide is found within	Construct(s) ID	Protein	Peptide ID.	Top MHC-I binding prediction
TVAYFNMVY	9	102, 201	t_cell_d3-mm11- 104, t_cell_d3-mc10- 1375	Orf1ab	sp4	HLA-A*01:01, HLA-A*03:01, HLA-A*29:02, HLA-B*15:01, HLA-C*03:03, HLA-C*07:01
VYYTSNPTTF	10	108, 197	t_cell_d3-mm11- 104, t_cell_d3-mc10- 1375	Orf1ab	sp5	HLA-C*07:01/02, HLA-A*23:01, HLA-A*29:02, HLA-C*04:01
LNKEMYLKL	9	24, 100, 188, 195	t_cell_d2-mc1000- 1717, t_cell_d3-	Orf1ab	sp10	HLA-B*08:01, HLA-C*07:01,

			mm11- 104, t_cell_d5- mm20- 620- wf, t_cell_d3- mc10- 1375			HLA- C*07:02
REVRTIKVF	9	108, 197	t_cell_d3- mm11- 104, t_cell_d3- mc10- 1375	Orf1ab	sp13	HLA- A*23:01, HLA- B*15:01/02, HLA- B*40:01, HLA- B*44:02, C*07:01, HLA- C*07:02
FHLDGEVITF	10	108, 197	t_cell_d3- mm11- 104, t_cell_d3- mc10- 1375	Orf1ab	sp15	HLA- A*01:01, HLA- A*02:01, HLA- A*23:01, HLA- B*15:01/02, HLA- C*03:03, HLA- C*04:01,

						HLA- C*07:01/02
ALYNKYKYF	9	100, 195	t_cell_d3- mm11- 104, t_cell_d3- mc10- 1375	Orf1ab	sp24	HLA- A*01:01, HLA- A*02:01, HLA- A*23:01, HLA- A*29:02, HLA- B*08:01, HLA- B*15:01/02, HLA- C*04:01, HLA- C*07:01/02
RARTVAGVSI	10	116, 152, 161, 180, 217	t_cell_d4- mm11- 193- wf, t_cell_d5- mc10-977, t_cell_d5- mm20- 620- wf, t_cell_d4- mc20-393	Orf1ab	sp25	HLA- B*07:02, HLA- B*08:01, HLA- B*15:01, HLA- C*03:03, HLA- C*07:01/02

TQYNRYLAL	9	100, 195	t_cell_d3- mm11- 104, t_cell_d3- mc10- 1375	Orf1ab	sp30	HLA- A*02:01, HLA- B*07:02, HLA- B*08:01, HLA- B*15:01/02, HLA- B*40:01, HLA- C*03:03, HLA- C*04:01, HLA- C*07:01/02
MNLKYAISA	9	116, 152, 161, 180, 217	t_cell_d4- mm11- 193- wf, t_cell_d5- mc10-977, t_cell_d5- mm20- 620- wf, t_cell_d4- mc20-393	Orf1ab	sp38	HLA- B*08:01, HLA- C*03:03
TQYNRYLALY	10	100, 195	t_cell_d3- mm11- 104,	Orf1ab	sp39	HLA- A*01:01, HLA- A*03:01,

			t_cell_d3- mc10- 1375			HLA- A*23:01, HLA- A*29:02, HLA- B*08:01, HLA- B*15:01/02, C*07:01/02.
NMVYMPASW	9	102, 201	t_cell_d3- mm11- 104, t_cell_d3- mc10- 1375	Orf1ab	sp41	HLA- A*23:01, HLA- A*29:02, HLA- B*15:01/02, HLA- C*03:03, HLA- C*07:02
KHFYWFFSNY	10	100, 195	t_cell_d3- mm11- 104, t_cell_d3- mc10- 1375	Orf1ab	sp42	HLA- A*23:01, HLA- A*29:02, C*07:01/02
KFYGGWHNM	9	182, 243	t_cell_d5- mm20- 620- wf, t_cell_d4- mc20-393	Orf1ab	sp43	HLA- A*02:01, HLA- A*23:01, HLA-

						A*29:02, HLA- C*03:03, HLA- B*15:01/02, HLA- C*04:01, HLA- C*07:01/02
TQMNLKYAI	9	116, 152, 161, 180, 217	t_cell_d4- mm11- 193- wf, t_cell_d5- mc10-977, t_cell_d5- mm20- 620- wf, t_cell_d4- mc20-393	Orf1ab	sp48	HLA- A*02:01, HLA- A*23:01, HLA- B*08:01, HLA- B*15:01/02, HLA- B*40:01, HLA- C*03:03, C*07:01/02
VMFTPLVPF	9	100, 195	t_cell_d3- mm11- 104, t_cell_d3- mc10- 1375	Orf1ab	sp49	HLA- A*02:01, HLA- A*23:01, HLA- A*29:02, HLA- B*15:01/02,

						HLA- C*03:03, HLA- C*07:02
STKHFYWFF	9	100, 195	t_cell_d3- mm11- 104, t_cell_d3- mc10- 1375	Orf1ab	sp50	HLA- A*23:01, HLA- A*29:02, HLA- B*15:01/02, HLA- C*07:02
MPASWVMRIM	10	51, 102, 201	t_cell_d2- mc1000- 1717, t_cell_d3- mm11- 104, t_cell_d3- mc10- 1375	Orf1ab	sp52	HLA- B*07:02, HLA- B*15:02, HLA- C*03:03
SWVMRIMTW	9	51, 102, 201	t_cell_d2- mc1000- 1717, t_cell_d3- mm11- 104, t_cell_d3- mc10- 1375	Orf1ab	sp56	HLA- A*23:01, HLA- A*29:02, HLA- B*08:01, HLA- B*44:02, HLA- C*04:01,

						HLA- C*07:01/02
YMPASWVMRI	10	51, 102, 201	t_cell_d2- mc1000- 1717, t_cell_d3- mm11- 104, t_cell_d3- mc10- 1375	Orf1ab	sp59	HLA- A*02:01, HLA- A*23:01, HLA- C*04:01, HLA- B*07:02
HFYWFFSNY	9	100, 195	t_cell_d3- mm11- 104, t_cell_d3- mc10- 1375	Orf1ab	sp62	HLA- A*01:01, HLA- A*03:01, HLA- A*23:01, HLA- A*29:02, HLA- C*04:01, HLA- C*07:01/02
RAAEIRASA	9	113, 203	t_cell_d3- mm11- 104, t_cell_d3- mc10- 1375	S	sp74	HLA- B*07:02, HLA- B*08:01, HLA- B*15:01/02, HLA-

						C*03:03, HLA- C*07:01
SNIIRGWIF	9	205	t_cell_d3- mc10- 1375	S	sp81	HLA- A*23:01, HLA- A*29:02, HLA- B*08:01, HLA- B*15:01/02, HLA- C*07:01/02
VYYPDKVFRS	10	205	t_cell_d3- mc10- 1375	S	sp93	HLA- A*23:01, HLA- A*29:02, HLA- C*07:02
RGWIFGTTL	9	205	t_cell_d3- mc10- 1375	S	sp106	HLA- A*23:01, HLA- B*07:02, HLA- B*08:01, HLA- C*03:03, HLA- C*04:01,

						HLA- C*07:01/02
--	--	--	--	--	--	--------------------

Sequences

The amino acid sequences of the present invention are defined as below:

SEQ ID NO	Amino acid sequence
1	FVDGVPFVVSIGYHYKELGIVMNMDVDTHRYRLSLKDLLYAADP
2	GPGPG
3	NRATLQAIASEFSSLPSYAAYATAQEAY
4	GGGS
5	QRFNNYMPYVFTLLFQLCTFTKSTNSRIRASLPTTIAKN
6	GP
7	AEEHFIETVSLAGSYKDWSSYSGQRTELGVEFLKRGDKIVYHTVGNVIEFHMEGE VLPLEKLKTLALREVKTIKVFTTVDNINLHTQVIDMSMTYGQQLG
8	GPGPG
9	LYRLVVDAYQAPPTI
10	GGGS
11	GFSKFVTVQGPPGTGKSHFAIGLAIYYPTARVVY
12	GPGPG
13	LHVTYVPSQERNFTTAPAICHEGKAYFPREGVFVF
14	GPGPG
15	IMHVSFLVMFSGVVPLWWTVLVYVVVVLGRHVFWFTSLCTKRTVHVGDLVFHSFQ DAALQTFMLDKEVFLRLKREISPETYARYLALFNKYKYSGPMDTA
16	GP
17	TSFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVNLHSSRLSFKELLVYAAD PAMHAASGNLLLDKRTTCFSVAALTN
18	GGGS
19	GPSDASALLYELAKRLEALEAGKSKSAPKVITKKDAADAKKKMRHKRVATKAY NPTQAFGLRGP
20	GP
21	ARLAQMCYAGVTLDVMRINKTHYSNLIGNVPDLFTKYNYALPTNFYGCVHAYYIN VTDGRYALATHYPATVITPGGRQPYNSYVATVLNTV
22	GFSKFVTVQGPPGTGKSHFAIGLAIYYPTARVVY
23	GP
24	TFSTFEEAALCTFLLNKEMYLKLRS
25	GPGPG

26	NVSVSRFNPSTWNRRFGFTEQSVFKPQP
27	GP
28	IVNVSLVKPTVYVYSRVKN
29	GP
30	ATIPLQASLPFGWLIGVAFLAVFQSATKIIALNKRWQLALYKGFQFI
31	GETALALLLDRLNQL
32	GGGS
33	VPIGGTTVVRPLVED
34	GPGPG
35	KINQALHGANLRQDESVYSLY
36	GPGPG
37	FAYSNRRNFLYIIKLVL
38	GP
39	LHVTYVPSQERNFTTAPAICHEGKAYFPREGVFVF
40	GP
41	LSPRWYFYLLGTGPEA
42	GGGS
43	QRFVNNYMPYVFTLLFQLCTFTKSTNSRIRASLPTTIKN
44	GGGS
45	ADLYKLMGHFSWWTAFVT
46	FAGTHITMTTVYHITVSQIQLSLLKVT
47	GGGS
48	RVAGSITMRSGSSPAL
49	AAIHYLNSKIVDLAQHFS
50	GP
51	VYMPASWVMRIMTWLEL
52	GGGS
53	ECTYISRFSKISLVSV
54	GGGS
55	LSVIVQPTHYVRVTFSDPNMWYLR
56	GPGPG
57	LPKEITVATSRTLSTYYKLG
58	GPGPG
59	ASSLFAQFASVTVNVTTAMGKALGR
60	GPGPG
61	PNNTASWFTALTQHGKEELRFPR
62	GPGPG
63	QFPILAEAPTAGAFFFGSRLELAKVQNLSG
64	QLGNSFGAIISSSINEIFSRLD
65	GPGPG
66	KYFKNHTSPDVLGDIS
67	FVDGVFPVVSIGYHYKELGIVMNMDVDTHRYRLSLKDLLLLYAADP
68	GP
69	HYYSDFIKIGSRKSARL
70	GGGS
71	NLDVLISSIVRQLFK

72	GGGS
73	NLDMISSIVRQLFK
74	GGGS
75	SELVIGAVIIRGHLRMAG
76	GPGPG
77	KIQTAPHISLAMLDIQPE
78	GGGS
79	GPMLRLIHLVSLWTF
80	GGGS
81	VVPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATE
82	GP
83	KSAGFPFNKFGKARLYYESLSY
84	GPGPG
85	ALTTFMITKDSYCKLKNSLSDVAFNRYL
86	GPGPG
87	RQKRTATKQYNVTQAFGRR
88	GPGPG
89	NRATLQAIASEFSSLPSYAAYATAQEAY
90	GP
91	GVAFLAVFQSATKIIALNKRWQLAL
92	GPGPG
93	QNKGLFYTQVYKNMAVYRSLTF
94	QGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNF KDQVILLNKHIDAYKTFPTE
95	GP
96	TGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQVAKSHNIALIWNVKDFM SLSEQLRKQIRSAACKNNLPFKLTCATTRQVVNVVTTKIALKGSKI
97	GPGPG
98	LFSTVFPLTSFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVNLHSSRLSFKE LLVYAADPAMHAASGNLLDKRTTCFSVAALTNNVAFQTVKPGNF
99	GPGPG
100	NDVSFLAHIQWMMFTPLVPFWITAIYIICISTKHFWFFSNYLKRRVVFNGVSFS TFEEAALCTFLLNKEMYKLKLRSDVLLPFTQYNRYLALYNKYKYF
101	GP
102	LPSLATVAYFNMVYMPASWWMRIMTWLDMVDTSLKLKDCVMYASAVVLLILMTA RTVYDDGARRVWTLMNVLTVYKVYYGNALDQAISMWALIISVTSN
103	GP
104	RATLQAIASEFSSLPSYA AFATAQEAYEQAVANGDSEVVLLKKLKKSLNVAKSEFD RDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTML
105	GGGS
106	RHINAQVAKSHNIALIWNVKDFMSLSEQLRKQIRSAACKNNLPFKLTCATTRQVV NVVTTKIALKGKIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVM
107	GP
108	EHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVITF DNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTY
109	GGGS
110	AVVYRGTTTYKLVGDYFVLTSHTVMPLSAPTLVPQEHYVRITGLYPTLNISDEF SSNVANYQKVGMMQKYSTLQGGPGTGKSHFAIGLALYYPSARIVYT

111	GP
112	LPKEITVATSRTLSYYKLG
113	AIGKIQDSLSTASALGKLQDVVNHNAQALNTLVKQLSSKFGAISSVLNDILSRDL KVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKM
114	EVVGNVILKPSDEGVKVTQELGHEDLMAAYVENTSITIKKPNELSLALGLKTIATH G
115	GP
116	DRYFEIYDGGCLNASEVIVNNLDKSAGHPFNKFGKARVYYESLSYQEQDEL FAM TKRNVLP TITQMNLYAISAKNRARTVAGVSILSTMTNRQYHQKMLKSMA
117	GPGPG
118	YNSDYASKGYIANISAFQQVLYYQNNVFMSEAKCWWENDINHGPHEFCSQHTM LVKMDGDDVYLPYPDP SRILGAGCFVDDLKTDSVLLIERFVSLAIDAY
119	GP
120	EYKGPVTDV FYKETS YTTTIKPVSYKLDGVTYTEIEPKLDGYYKKDNAYVENTSIT IKKPNELSLALGLKTIATHGIAAINSVPWSKILAYVKPFLGQAA
121	GP
122	AEHFIE TVSLAGSYKDWSYSGRTELGV EFLKRGDKIVYHTVGNVIEFHMEGE VLPLEKLKTL LALREVKTIKVFTTVDNINLHTQVIDMSMTYGQQLG
123	GPGPG
124	NARDGCVPLNIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQQVDA DSKIVQLSEINMDNSQNLAWPLIVTALRANSVKLQNNELSPVALR
125	GGGS
126	PIEFHLDGEVLP LDKLSLLSLREVKTIKVFTTVDNTNLHTQLVDM SVTYGQHFG PTY
127	GP
128	KMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLNDALNNIINNARDG CVPLNIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQQVV
129	GPGPG
130	VSLIWNVKDYMSLSEQLRKQIRSAKKNNIPFRLTCATTRQVV
131	GP
132	LAYMLFTKFFYLLGLSAIMQVFFGYFASHFISN
133	GGGS
134	VLKKLKKSLNVAKSEFDRDAAMQRKLEKMADQAMTQMYKQARSEDKRAKVTS AMQTMLFTML
135	KMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLNDALNNIINNARDG CVPLNIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQQVV
136	GPGPG
137	HHANEYRQYLDAYNMMISAGFSLWIYKQFDTYNLWNTFTRLQSL ENVAYNVVN KGHFDGH
138	GP
139	LHVTYVPSQERNFTTAPAICHEGKAYFPREGVFVF
140	GP
141	TSFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVNLHSSRLSFKELLVYAAD PAMHAASGNLLLDKRTTCFSVAALTN
142	GP
143	VFISPYNSQNAVASKILGLPTQTVDSSQGSEYDYVIFTQTTETAHSCNVNRFNVA ITRAKIGILCIMS DRDLYDKLQFTSLEIPRRNVATLQAENV TGLFKD
144	GGGS
145	EYKGPVTDV FYKETS YTTTIKPVSYKLDGVTYTEIEPKLDGYYKKDNAYVENTSIT IKKPNELSLALGLKTIATHGIAAINSVPWSKILAYVKPFLGQAA

146	GGGS
147	NARDGCVPLNIIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQVVDADSKIVQLSEINMDNSQNLAWPLIVTALRANSVAVKLQNNELSPVALR
148	GPGPG
149	AEHFIEFVSLAGSYKDWSYSGQRTELGVFLKRGDKIVYHTVGNVIEFHMEGEVLPLEKLTLLALREVKTIVFTTVDNINLHTQVIDMSMTYGQQLG
150	YNSDYASKGYIANISAFQQVLYYQNNVFMSEAKCWWENDINHGPHEFCSQHTMLVKMDGDDVYLPYPDPSRILGAGCFVDDLLKTDSVLLIERFVSLAIDAY
151	GPGPG
152	DRYFEIYDGGCLNASEVIVNNLDKSAGHPFNKFGKARVYYESLSYQEQDELFTAMTKRNVLPITITQMNLYAISAKNRARTVAGVSILSTMTNRQYHQKMLKSMA
153	PIEFHLDGEVLPDLKLSLLSLREVKTIVFTTVDNTNLHTQLVDMSVTYGQHFGPTY
154	GP
155	AEHFIEFVSLAGSYKDWSYSGQRTELGVFLKRGDKIVYHTVGNVIEFHMEGEVLPLEKLTLLALREVKTIVFTTVDNINLHTQVIDMSMTYGQQLG
156	GP
157	HHANEYRQYLDAYNMMISAGFSLWIYKQFDTYNLWNTFTRLQSLNVAYNVNVNKGHFDGH
158	GP
159	LAYMLFTKFFYLLGLSAIMQVFFGYFASHFISN
160	GGGS
161	TKRNVLPITITQMNLYAISAKNRARTVAGVSI
162	GP
163	YNSDYASKGYIANISAFQQVLYYQNNVFMSEAKCWWENDINHGPHEFCSQHTMLVKMDGDDVYLPYPDPSRILGAGCFVDDLLKTDSVLLIERFVSLAIDAY
164	GP
165	GLDSLDSYPALETIQVTISSYKDLT
166	GP
167	ALTTFMITKDSYCKLKNSLSDVAFNRYL
168	GPGPG
169	PAYSFLPGVYSVFYLYLTFYFTNDVSFLAHLQWFAMFSPIVPFWIT
170	GPGPG
171	NARDGCVPLNIIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQVVDADSKIVQLSEINMDNSQNLAWPLIVTALRANSVAVKLQNNELSPVALREVVGNVILKPSDEGVKTQELGHEDLMAAYVENTSITIKKPNELSLALGLKTIATHG
172	G
173	GGGS
174	TSFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVNLHSSRSLFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTN
175	GPGPG
176	EKGPFHEFCSQHTMLVKMDGDEVYLPYPDPSRILGAGCFVDDLLKTDSVLLIERFVSLAIDAYPLVHHENPEYQNVFRVYLEYIKKLYNDLGNQILDSYSV
177	GPGPG
178	KMADQAMTQMYKQARSEDKRAKVTSAMQTMFTMLRKLDNDALNNIINNARDGCVPLNIIPLTTAAKLMVVVPDYGTYNKTCGNTFTYASALWEIQVVD
179	GPGPG
180	DRYFEIYDGGCLNASEVIVNNLDKSAGHPFNKFGKARVYYESLSYQEQDELFTAMTKRNVLPITITQMNLYAISAKNRARTVAGVSILSTMTNRQYHQKMLKSMA

181	GGGS
182	RTVAGVSICTMTNRQFHQKLLKSIAATRGATVWIGTSKFYGGWHNMLKTVYSD
183	VE
184	GPGPG
185	EYKGPVTDVIFYKETSYYTTTIKPVSYKLDGVTYTEIEPKLDGYYKKNAYVENTSIT
186	IKKPNELSLALGLKTIATHGIAAINSVPWSKILAYVKPFLGQAA
187	GP
188	VSLIWNVKDYMSLSEQLRKQIRSAKKNNIPFRLTCATTRQVV
189	GGGS
190	TFSTFEEAALCTFLLNKEMYLKLRSE
191	GPGPG
192	VLKLLKSLNVAKSEFDRDAAMQRKLEKMAQAMTQMYQARSEDKRAKVTS
193	AMQTMLFTML
194	RATLQAIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKLLKSLNVAKSEFD
195	RDAAMQRKLEKMAQAMTQMYQARSEDKRAKVTSAMQTMLFTML
196	GP
197	AVVYRGTTTTYKLVNGDYFVLTSHTVMPLSAPTLVPQEHYVRITGLYPTLNISDEF
198	SSNVANYQKVGMMQKYSTLQGPPGTGKSHFAIGLALYYPSARIVYT
199	GPGPG
200	NDVSFLAHIQWMMFTPLVPFWITAIYIICISTKHFYWFFSNYLKRRRVFNGVSFS
201	TFEEAALCTFLLNKEMYLKLRSDVLLPFTQYNRYLALYNKYKYF
202	GPGPG
203	EHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDKSVYYTSNPTTFHLDGEVITF
204	DNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTY
205	GPGPG
206	LFSTVFPLTSFGPLVRKIFVDGVFPVSTGYHFRELGVVHNQDVNLHSSRLSFKE
207	LLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNF
208	GP
209	LPSLATVAYFNMVYMPASWWMRIMTWLDMVDTSLKLKDCVMYASAVLLILMTA
210	RTVYDDGARRVWTLMNVLTVYKVYYGNALDQAISMWALIISVTSN
211	GPGPG
212	AIGKIQDSLSTASALGKLQDVVNHNAAQALNTLVKQLSSKFGAISSVLNDILSRD
213	KVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKM
214	GP
215	QSYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFWHAIHVSGTNGTKRF
216	DNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSK
217	LPKEITVATSRTLSTYYKLG
218	GGGS
219	RHINAQVAKSHNIALIWNVKDFMSLSEQLRKQIRSAKKNNLPFKLTCATTRQVV
220	NVVTTKIALKGKIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVM
221	DVLISIVRQLFKPL
222	GGGS
223	LHVTYVPSQERNFTTAPAICHEGKAYFPREGVFVF
224	GPGPG
225	AEHFIEITVSLAGSYKDWSYSGQRTELGVFLKRGDKIVYHTVGNVIEFHMEGE
226	VLPLEKLTLLALREVKTIKVFTTVDNINLHTQVIDMSMTYGQQLG
227	GPGPG
228	LPKEITVATSRTLSTYYKLG
229	GPGPG

217	DRYFEIYDGGCLNASEVIVNNLDKSAGHPFNKFGKARVYYESLSYQEQDEL FAM TKRNVLP TITQMN LKYAISAKNRARTVAGVSILSTMTNRQYHQKMLKSMA
218	LAYMLFTKFFYLLGLSAIMQVFFGYFASHFISN
219	SVLQATLSEFSLASYAELEAAQSSYQKALNSGDASPQVLKALQKAVNIAKNAY EKDKAVARKLERMAEQAMTS MYQARAEDKKAKIVSAMQTM
220	GGGS
221	KMADQAMTQMYQARSEDKRAKVT SAMQTM LFTMLRKL DNDALNNIINNARDG CVPLNIPLTTAAKLMVVVPDYGT YKNTCDGNTFTYASALWEIQQVV
222	GGGS
223	HHANEYRQYLDAYNMMISAGFSLWIYKQFDTYNLWNTFTRLQSL ENVAYNVVN KGHFDGH
224	GPGPG
225	NRATLQAIASEFSSLPSYAA YATAQEAY
226	GPGPG
227	YNSDYASKGYIANISAFQQVLYYQNNVFMSEAKCWWENDINHGPHEFCSQHTM LVKMDGDDVYLPYPDPSRILGAGCFVDDLLKTDSVLLIERFVSLAIDAY
228	GP
229	TSLEIPRRNVATLQAENV TGLFKDCSKIITGLHPTQAPTHLSVDIKFKTEGLCVDIP GIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRAW
230	GGGS
231	GSNLEEEILLMLASLFGTYTWTTVLSMA
232	GGGS
233	LSYFVASFRLFARTRS
234	GPGPG
235	KYFKNHTSPDVLGDIS
236	GGGS
237	FVDGVPFVVSIGYHYKELGIVMNM DVDTHRYRLSLKDLLLYAADP
238	GPGPG
239	EYKGPVTDV FYKETS YTTTIKPVSYKLDGVTYTEIEPKLDGYYKKDNAYVENTSIT IKKPNELSLALGLKTIATHGIAAINSVPWSKILAYVKPFLGQAA
240	GP
241	SSYMAISLKFTTSLSLHKLLQTLVLK
242	GPGPG
243	RTVAGVSICSTMTNRQFHQKLLKSIAATRGATV VIGTSKFYGGWHNMLKTVYSD VE
244	GP
245	VFISPYNSQNAVASKILGLPTQTVDSSQGSEYDYVIFTQTTETAHSCNVNRFNVA ITRAKIGILCIMS DRDLYDKLQFTSLEIPRRNVATLQAENV TGLFKD
246	FVDGVPFVV
247	AIASEFSSL
248	FSSLPSYAA
249	SEFSSLPSY
250	TLQAIASEF
251	IFVDGVPFV
252	KIFVDGVPF
253	LLDKRTTCF
254	LVYAADPAM
255	ASGNLLLDK

256	LSFKELLY
257	RLSFKELLV
258	TSFGPLVRK
259	VVSTGYHFR
260	FVVSTGYHF
261	NLHSSRLSF
262	SSRLSFKEL
263	VPFVVSTGY
264	FVDGVPFVV
265	LLDRLNQL
266	VATSRTLSTY
267	ATSRTLSTYY
268	LPKEITVAT
269	TSRTLSTYYK
270	LSPRWYFYFYY
271	SPRWYFYFYYL
272	YYLGTGPEA
273	AIASEFSSL
274	FSSLPSYAA
275	SEFSSLPSY
276	TLQAIASEF
277	FTALTQHGK
278	SELVIGAVI
279	ALCTFLLNK
280	FLLNKEMYL
281	LLNKEMYLK
282	EEAALCTFL
283	FEEAALCTF
284	LNKEMYLKL
285	TFLLNKEMY
286	MPASWWMRI
287	SWWMRIMTW
288	VYMPASWWM
289	WWMRIMTWL
290	VATSRTLSTY
291	ATSRTLSTYY
292	LPKEITVAT
293	TSRTLSTYYK
294	AMQTMLFTM
295	MQTMLFTML
296	KRAKVTSAM
297	KVTSAMQTM
298	MADQAMTQM
299	TSAMQTMLF
300	AAMQRKLEK
301	LEKMADQAM

302	SEFDRDAAM
303	SLNVAKSEF
304	VLKKLKKSL
305	KQIRSAAKK
306	MSLSEQLRK
307	QLRKQIRSA
308	AMQTMLFTM
309	MQTMLFTML
310	KRAKVTSAM
311	KVTSAMQTM
312	MADQAMTQM
313	TSAMQTMLF
314	FVDGVPFVV
315	IFVDGVPFV
316	KIFVDGVPF
317	LLDKRTTCF
318	LVYAADPAM
319	ASGNLLLDK
320	LSFKELLY
321	RLSFKELLV
322	TSFGPLVRK
323	VVSTGYHFR
324	FVVSTGYHF
325	NLHSSRLSF
326	SSRLSFKEL
327	VPFVVSTGY
328	AMQTMLFTM
329	MQTMLFTML
330	KRAKVTSAM
331	KVTSAMQTM
332	MADQAMTQM
333	TSAMQTMLF
334	ALCTFLLNK
335	FLLNKEMYL
336	LLNKEMYLK
337	EEAALCTFL
338	FEEAALCTF
339	LNKEMYLKL
340	TFLLNKEMY
341	FVDGVPFVV
342	IFVDGVPFV
343	KIFVDGVPF
344	LLDKRTTCF
345	LVYAADPAM
346	ASGNLLLDK
347	LSFKELLY

348	RLSFKELLV
349	TSFGPLVRK
350	VVSTGYHFR
351	FVVSTGYHF
352	NLHSSRLSF
353	SSRLSFKEL
354	VPFVVSTGY
355	AAMQRKLEK
356	LEKMADQAM
357	SEFDRDAAM
358	SLNVAKSEF
359	VLKCLKKSL
360	KQIRSAKK
361	MSLSEQLRK
362	QLRKQIRSA
363	TVAYFNMVY
364	VYYTSNPTTF
365	LNKEMYLKL
366	REVRTIKVF
367	FHLDGEVITF
368	ALYNKYKYF
369	RARTVAGVSI
370	TQYNRYLAL
371	MNLKYAISA
372	TQYNRYLALY
373	NMVYMPASW
374	KHFYWFFSNY
375	KFYGGWHNM
376	TQMNLKYAI
377	VMFTPLVPF
378	STKHFWFF
379	MPASWVMRIM
380	SWVMRIMTW
381	YMPASWMRI
382	HFYWFFSNY
383	RAAEIRASA
384	SNIIRGWIF
385	VYYPDKVFRS
386	RGWIFGTTL

Embodiments

1. A polypeptide comprising one or more epitope sequences, wherein the one or more epitope sequences have the amino acid sequences of any one or

more of SEQ ID Nos 1 to 245, or wherein the one or more epitope sequences are variants of the sequences of any one or more of SEQ ID Nos 1 to 245 wherein the variant sequences have at least 70% sequence identity to any one or more of SEQ ID Nos 1 to 245, and wherein the polypeptide sequence is no more than 1400
5 amino acids in length.

2. The polypeptide according to embodiment 1, comprising at least 2, 3, 4, 5, 6, 7, 8, 9, 10 or more epitope sequences having the amino acid sequences of any of SEQ ID Nos 1 to 245, or the variants thereof.

3. The polypeptide according to embodiments 1 or 2, comprising a
10 concatemer of epitope sequences having the amino acid sequences of any one or more of SEQ ID Nos 1 to 245, or the variants thereof.

4. The polypeptide according to any of embodiments 1 to 3, comprising each epitope in any one of the following groups of epitope sequences:

- a. SEQ ID No 1 to 21, or
- 15 b. SEQ ID No 22 to 93, or
- c. SEQ ID No 94 to 113, or
- d. SEQ ID No 114 to 134, or
- e. SEQ ID No 135 to 152, or
- f. SEQ ID No 153 to 190, or
- 20 g. SEQ ID No 191 to 208, or
- h. SEQ ID No 209 to 245.

5. The polypeptide according to any of embodiments 1 to 4, comprising each epitope in any of the following groups of epitope sequences arranged in the following order on the polypeptide:

- a. SEQ ID No 1 to 21, or
 - b. SEQ ID No 22 to 93, or
 - c. SEQ ID No 94 to 113, or
 - d. SEQ ID No 114 to 134, or
 - 5 e. SEQ ID No 135 to 152, or
 - f. SEQ ID No 153 to 190, or
 - g. SEQ ID No 191 to 208, or
 - h. SEQ ID No 209 to 245.
6. The polypeptide according to any of embodiments 1 to 5, wherein the
10 polypeptide comprises one or more spacer regions; preferably wherein when the polypeptide comprises more than one of the epitope sequences of SEQ ID Nos 1 to 245 or the variants thereof, one or more spacer regions are positioned between one or more or each epitope sequence.
7. The polypeptide according to any of embodiments 1 to 6, wherein each
15 of the one or more epitope sequences have at least 70%, 80%, 90%, 95%, 96%, 97%, 98% or 99% sequence identity to any of SEQ ID Nos 1 to 245.
8. The polypeptide according to any of embodiments 1 to 7, wherein said
polypeptide is no more than 1400, 1300, 1200, 1100, 1000, 900, 800, 700, 600,
500, 400, 300, 200, 150, 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 10 or 9 amino
20 acids in length.
9. A polynucleotide encoding a polypeptide as defined in any of
embodiments 1 to 8.
10. The polynucleotide according to embodiment 9, wherein the
polynucleotide is DNA.

11. The polynucleotide according to embodiment 9, wherein the polynucleotide is RNA, preferably mRNA.
12. A vector comprising a polynucleotide according to any of embodiments 9 to 11, preferably wherein the vector further comprises regulatory elements capable of driving transcription and/or translation of the polynucleotide in a host cell.
13. The vector according to embodiment 12, wherein the vector is a viral vector.
14. A microorganism comprising a polypeptide according to any of embodiments 1 to 8, or a polynucleotide according to any of embodiments 9 to 11, preferably wherein said microorganism is a bacterial microorganism.
15. The polynucleotide according to any of embodiments 9 to 11 or a vector according to embodiment 12, formulated in a nanoparticle or lipid formulation.
16. A composition comprising:
- i) one or more epitope sequences comprising any one or more of SEQ ID Nos 1 to 245, or variants thereof having at least 70% sequence identity to any one or more of SEQ ID Nos 1 to 245, each of said epitope sequences or variants thereof being no more than 1400 amino acids in length, or
 - ii) one or more polynucleotides encoding one or more epitope sequences comprising any one or more of SEQ ID Nos 1 to 245, or variants thereof having at least 70% sequence identity to any one or more of SEQ ID Nos 1 to 245, each of said epitope sequences or variants thereof being no more than 1400 amino acids in length.
17. The polypeptide, polynucleotide, vector, microorganism or composition according to any preceding embodiment, for therapeutic or prophylactic use.

18. The polypeptide, polynucleotide, vector, microorganism or composition according to any of embodiments 1 to 17, for use in the treatment or prophylaxis of an infection caused by a Betacoronavirus.
19. The polypeptide, polynucleotide, vector, microorganism or composition
5 for use according to embodiment 18, wherein the infection is caused by a member of the Embecovirus, Hibecovirus, Merbecovirus, Nobecovirus or Sarbecovirus subgenera, or by MERS-CoV, SARS-CoV or SARS-CoV-2.
20. A vaccine composition comprising a polypeptide, polynucleotide, vector or microorganism according to any of embodiments 1 to 15.
- 10 21. The vaccine composition according to embodiment 20, wherein the vaccine is an mRNA vaccine.
22. The vaccine composition according to embodiment 20 or embodiment 21, further comprising a pharmaceutically acceptable carrier, diluent, excipient and/or adjuvant.
- 15 23. The vaccine composition according to any of embodiments 20 to 22, wherein said composition is formulated for a parental, oral, sublingual, nasal, naso-oral, or pulmonary administration.
24. The vaccine composition according to embodiment 23, wherein said parental administration is subcutaneous, intradermal, intramuscular, subdermal,
20 intraperitoneal, or intravenous administration.

CLAIMS

1. A vaccine composition comprising a polypeptide, wherein the polypeptide comprises one or more epitope sequences, wherein the one or more epitope sequences have the amino acid sequences of any one or more of SEQ ID Nos 1,
5 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 24, 26, 28, 30, 31, 33, 35, 37, 41, 45, 46, 48, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 91, 93, 94, 96, 98, 100, 102, 104, 106, 108, 110, 113, 114, 116, 118, 120, 124, 126, 128, 130, 132, 134, 137, 143, 161, 165, 169, 176, 182, 205, 209, 219, 229, 231, 233 and/or 241, or a variant thereof having at least 70% sequence identity thereto, and
10 wherein the polypeptide sequence is no more than 1400 amino acids in length.
2. The vaccine composition according to claim 1, wherein the polypeptide comprises at least 2, 3, 4, 5, 6, 7, 8, 9, 10 or more epitope sequences having the amino acid sequences of any of SEQ ID Nos 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 24, 26, 28, 30, 31, 33, 35, 37, 41, 45, 46, 48, 49, 51, 53, 55, 57, 59, 61, 63, 64,
15 66, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 91, 93, 94, 96, 98, 100, 102, 104, 106, 108, 110, 113, 114, 116, 118, 120, 124, 126, 128, 130, 132, 134, 137, 143, 161, 165, 169, 176, 182, 205, 209, 219, 229, 231, 233 and/or 241, or a variant thereof.
3. The vaccine composition according to claim 1 or 2, wherein the polypeptide comprises a concatemer of epitope sequences having the amino acid
20 sequences of any one or more of SEQ ID Nos 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 24, 26, 28, 30, 31, 33, 35, 37, 41, 45, 46, 48, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 91, 93, 94, 96, 98, 100, 102, 104, 106, 108, 110, 113, 114, 116, 118, 120, 124, 126, 128, 130, 132, 134, 137, 143, 161, 165, 169, 176, 182, 205, 209, 219, 229, 231, 233 and/or 241, or a variant thereof.
- 25 4. The vaccine composition according to any of claims 1 to 3, wherein the polypeptide comprises each epitope in any one of the following groups of epitope sequences:

a. SEQ ID No 1 to 21, or

- b. SEQ ID No 22 to 93, or
- c. SEQ ID No 94 to 113, or
- d. SEQ ID No 114 to 134, or
- e. SEQ ID No 135 to 152, or
- 5 f. SEQ ID No 153 to 190, or
- g. SEQ ID No 191 to 208, or
- h. SEQ ID No 209 to 245.

5. The vaccine composition according to any of claims 1 to 4, wherein the polypeptide comprises each epitope in any of the following groups of epitope
10 sequences arranged in the following order on the polypeptide:

- a. SEQ ID No 1 to 21, or
- b. SEQ ID No 22 to 93, or
- c. SEQ ID No 94 to 113, or
- d. SEQ ID No 114 to 134, or
- 15 e. SEQ ID No 135 to 152, or
- f. SEQ ID No 153 to 190, or
- g. SEQ ID No 191 to 208, or
- h. SEQ ID No 209 to 245.

6. The vaccine composition according to any of claims 1 to 5, wherein the
20 polypeptide comprises one or more spacer regions; preferably wherein when the polypeptide comprises more than one of the epitope sequences of SEQ ID Nos 1,

3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 24, 26, 28, 30, 31, 33, 35, 37, 41, 45, 46, 48, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 91, 93, 94, 96, 98, 100, 102, 104, 106, 108, 110, 113, 114, 116, 118, 120, 124, 126, 128, 130, 132, 134, 137, 143, 161, 165, 169, 176, 182, 205, 209, 219, 229, 231, 233
5 and/or 241 or a variant thereof, one or more spacer regions are positioned between one or more or each epitope sequence.

7. The vaccine composition according to any of claims 1 to 6, wherein each of the one or more epitope sequences have at least 70%, 80%, 90%, 95%, 96%, 97%, 98% or 99% sequence identity to any of the epitope sequences as defined
10 in any of claims 1 to 6.

8. The vaccine composition according to any of claims 1 to 7, wherein said polypeptide is no more than 1400, 1300, 1200, 1100, 1000, 900, 800, 700, 600, 500, 400, 300, 200, 150, 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 10 or 9 amino acids in length.

15 9. The vaccine composition according to any of claims 1 to 8, wherein the vaccine is an mRNA vaccine.

10. The vaccine composition according to any of claims 1 to 9, further comprising a pharmaceutically acceptable carrier, diluent, excipient and/or adjuvant.

20 11. The vaccine composition according to any of claims 1 to 10, wherein said composition is formulated for a parental, oral, sublingual, nasal, naso-oral, or pulmonary administration.

12. The vaccine composition according to claim 11, wherein said parental administration is subcutaneous, intradermal, intramuscular, subdermal,
25 intraperitoneal, or intravenous administration.

13. A vaccine composition according to any of claims 1 to 12, for therapeutic or prophylactic use.

14. A vaccine composition according to any of claims 1 to 12, for use in the treatment or prophylaxis of an infection caused by a Betacoronavirus.
15. The vaccine composition according to claim 13, wherein the infection is caused by a member of the Embecovirus, Hibecovirus, Merbecovirus, Nobecovirus or Sarbecovirus subgenera, or by MERS-CoV, SARS-CoV or SARS-CoV-2.
16. Use of a vaccine composition according to any of claims 1 to 12 in the manufacture of a medicament for the treatment of a Betacoronavirus infection.
17. A method of treating a Betacoronavirus infection, comprising administering to a subject an effective amount of the vaccine composition according to any of claims 1 to 12.
18. A vector comprising a polynucleotide encoding the polypeptide as defined in any of claims 1 to 5, preferably wherein the vector further comprises regulatory elements capable of driving transcription and/or translation of the polynucleotide in a host cell.
19. The vector according to claim 18, wherein the vector is a viral vector.
20. A microorganism comprising a polypeptide as defined in any of claims 1 to 5, or a polynucleotide encoding the polypeptide as defined in any of claims 1 to 5, preferably wherein said microorganism is a bacterial microorganism.
21. The polynucleotide encoding the polypeptide as defined in any of claims 1 to 5, or a vector according to claims 18 or 19, formulated in a nanoparticle or lipid formulation.
22. A composition comprising:
- i) one or more epitope sequences comprising any one or more of SEQ ID Nos 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 24, 26, 28, 30, 31, 33, 35, 37, 41, 45, 46, 48, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87,

91, 93, 94, 96, 98, 100, 102, 104, 106, 108, 110, 113, 114, 116, 118, 120, 124, 126, 128, 130, 132, 134, 137, 143, 161, 165, 169, 176, 182, 205, 209, 219, 229, 231, 233 and/or 241, or a variant thereof having at least 70% sequence identity, each of said epitope sequences or variants thereof being no more than 1400 amino acids in length, or

ii) one or more polynucleotides encoding one or more epitope sequences comprising any one or more of SEQ ID Nos 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 24, 26, 28, 30, 31, 33, 35, 37, 41, 45, 46, 48, 49, 51, 53, 55, 57, 59, 61, 63, 64, 66, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 91, 93, 94, 96, 98, 100, 102, 104, 106, 108, 110, 113, 114, 116, 118, 120, 124, 126, 128, 130, 132, 134, 137, 143, 161, 165, 169, 176, 182, 205, 209, 219, 229, 231, 233 and/or 241, or a variant thereof having at least 70% sequence identity, each of said epitope sequences or variants thereof being no more than 1400 amino acids in length.

23. The vector, microorganism, polynucleotide or composition according to any of claims 18 to 22, for therapeutic or prophylactic use.

24. The vector, microorganism, polynucleotide or composition according to claim 23, for use in the treatment or prophylaxis of an infection caused by a Betacoronavirus.

25. The vector, microorganism, polynucleotide or composition for use according to claim 24, wherein the infection is caused by a member of the Embecovirus, Hibecovirus, Merbecovirus, Nobecovirus or Sarbecovirus subgenera, or by MERS-CoV, SARS-CoV or SARS-CoV-2.