

EPITOPES OF SARS-CoV-2

5 TECHNICAL FIELD

The present disclosure relates broadly to the identification of epitopes within Severe Acute Respiratory Syndrome (SARS) Viruses for use as diagnostic as well as prophylactic and therapeutic use.

10

BACKGROUND

The ongoing unprecedented coronavirus disease 2019 (COVID-19) pandemic is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Since its first explosive outbreak in Wuhan, China in December 2019, SARS-CoV-2 has now caused at least 100 million infection cases with 2.3 million deaths globally (as of 8 February 2021).

This unprecedented crisis has generated major socio-economic disruptions worldwide and overwhelmed healthcare systems in both developed and developing countries. While quantitative reverse transcriptase PCR (qRT-PCR) remains as a gold standard for the early detection of SARS-CoV-2-infected individuals, the incidence of false negative diagnosis due to insufficient viral genetic material at the point of detection has limited such molecular-based assays (3). Although rapid tests to detect acute infections are crucial in light of the current pandemic, it is also necessary to develop efficient assays for the detection of exposure in order to successfully mitigate virus transmission. Thus, there is an imperative need to rapidly develop highly specific and sensitive serology-based approaches to confirm molecular diagnosis, and enable immunosurveillance to identify individuals with recent or past exposure to SARS-CoV-2.

To date, several studies and pre-prints have described serological assays to detect SARS-CoV-2-specific antibodies in patient blood specimens (3-7). These immunoassays primarily use recombinant coronavirus proteins, such as

the spike (S), which is exposed on the surface of the viral particle, and the nucleocapsid (N), which is highly expressed during infection (5, 8). The wealth of findings from these studies has shed some light on the general humoral response upon SARS-CoV-2 infection, where anti-SARS-CoV-2-IgM and IgG are detectable as early as 3 and 4 days post-illness onset (pio), respectively (3, 7). However, the use of recombinant proteins may give rise to batch-to-batch variations, which would greatly affect the reproducibility of immunoassays.

In addition, given the seasonal circulation of several human coronaviruses (hCoV) including 229E, OC43, NL63 and HKU1, which are responsible for common cold (9), coupled with individuals who have prior exposure to Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) during the outbreak in 2003 (10), cross-reactivity amongst these genetically similar viruses is inevitable, which may lead to confounding detection outcomes (11). Thus, it remains crucial to ascertain the specificity and sensitivity profiles for the development of efficient and precise immunoassays.

Lastly, specific regions on the viral proteins where these virus-specific antibodies recognise have yet to be further characterised, which would be fundamental to allow for a more targeted immune-based approach.

There is a need to provide epitopes of SARS-CoV-2. There is a need to provide epitopes of SARS-CoV-2 for use in diagnostic, prophylactic and/or therapeutic uses.

SUMMARY

In one aspect, there is provided an isolated immunogenic polypeptide or epitope comprising at least 12 contiguous amino acids of one or more SARS-CoV-2 protein selected from the group consisting of a spike (S) protein, an envelope (E) protein, a membrane (M) protein, and a nucleocapsid (N) protein.

In some examples, the isolated immunogenic polypeptide or epitope comprises 18 to 24 contiguous amino acids of spike (S) protein and/or nucleocapsid (N) protein of SARS-CoV-2.

In some examples, the isolated immunogenic polypeptide or epitope is selected from a polypeptide comprising an amino acid sequence selected from

the group consisting of: SEQ ID NO: 1 to 5, 11 to 38, 39 to 48, 49, 50, and 51 to 60 or a functional variant or fragment thereof; a polypeptide sharing at least 75% sequence identity thereto; or a polypeptide differing by one, two, three, or four amino acids thereto; or portions thereof.

5 In some examples, the isolated immunogenic polypeptide or epitope comprises an amino acid sequence selected from the group consisting of: SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 55, and SEQ ID NO: 57; or a peptide sharing at least 75% sequence identity thereto; or a peptide differing by one, two,
10 three, or four amino acids thereto; or portions thereof.

 In some examples, the polypeptide or epitope induces an immune response, optionally the polypeptide or epitope induces a B cell immune response.

 In some examples, the polypeptide or epitope does not include a
15 polypeptide or epitope of seasonal coronavirus or SARS-CoV.

 In some examples, the polypeptide or epitope does not include a polypeptide or epitope of one or more SARS-CoV protein selected from the group consisting of a spike (S) protein, an envelope (E) protein, a membrane (M) protein, and a nucleocapsid (N) protein.

20 In another aspect, there is provided a pharmaceutical composition comprising one or more isolated polypeptide as disclosed herein and a pharmaceutically acceptable carrier.

 In some examples, the pharmaceutical composition further comprising an adjuvant.

25 In yet another aspect, there is provided a composition comprising one or more isolated polypeptide as disclosed herein.

 In yet another aspect, there is provided an antigen binding molecule capable of specifically binding to the polypeptide or epitope as disclosed herein.

 In some examples, the antigen binding molecule is one selected from the
30 group consisting of IgG, IgM, IgE, IgA, IgD, F(ab) fragments, and fragments thereof.

In yet another aspect, there is provided a method of detecting the presence of SARS-CoV-2 virus in a sample comprising contacting the sample to one or more polypeptide as disclosed herein and determining the binding of a component of the sample to the polypeptide, wherein the binding of a component of the sample to the polypeptide is indicative of the presence of SARS-CoV-2.

In yet another aspect, there is provided a method of predicting the severity of a SARS-CoV-2 infection comprising contacting the sample to one or more polypeptide selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 50, SEQ ID NO: 55, and SEQ ID NO: 57 and determining the binding of a component of the sample to the polypeptide, wherein the binding of a component of the sample to the polypeptide is indicative of an increased risk of a severe SARS-CoV-2 infection.

In some examples, the sample comprises a biological sample, optionally the sample is a fluid biological sample.

In yet another aspect, there is provided a kit comprising one or more polypeptide as disclosed herein, the antigen binding molecule as disclosed herein.

In some examples, the kit further comprising one or more of the following: a substrate coated with a capture agent for anti-SARS-CoV-2 antigen binding molecule, and a detection agent for detecting the presence of captured anti-SARS-CoV-2 antigen binding molecule, wherein the capture agent and the detection agent comprise a SARS-CoV-2 antigen and/or anti-immunoglobulin.

In some examples, the polypeptide is selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 55, and SEQ ID NO: 57.

In yet another aspect, there is provided a method of treating or preventing SARS-CoV-2 in a subject in need thereof comprising administering into the subject a polypeptide as disclosed herein.

In yet another aspect, there is provided a polypeptide as disclosed herein for use as a medicament.

DEFINITIONS

The term "epitope" includes any moiety that is specifically recognised by an antigen binding protein or molecule (for example an immunoglobulin, an antibody or a receptor) in whole or in part. In some examples, the epitope
5 comprises of a plurality of amino acids in an antigen that are physically next to one another when in linear formation (i.e. a linear epitope).

The term "SARS-CoV-2" includes a virus having the sequence set forth in NCBI GenBank accession number MN908947. In some examples, the term "SARS-CoV-2" includes a virus comprising the sequence set forth in SEQ ID NO:
10 69 (ORF1ab polyprotein), SEQ ID NO: 70 (surface glycoprotein), SEQ ID NO: 71 (ORF3a protein), SEQ ID NO: 72 (envelope protein), SEQ ID NO: 73 (membrane glycoprotein), SEQ ID NO: 74 (ORF6 protein), SEQ ID NO: 75 (ORF7a protein), SEQ ID NO: 76 (ORF8 protein), SEQ ID NO: 77 (nucleocapsid phosphoprotein), SEQ ID NO: 78 (ORF10 protein), or encoded by sequence comprising SEQ ID
15 NO: 79 (complete genome) as well as variants/mutants thereof that share at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least
20 about 96%, at least about 97%, at least about 98%, at least about 99%, at least about 99.1%, at least about 99.2%, at least about 99.3%, at least about 99.4%, at least about 99.5%, at least about 99.6%, at least about 99.7%, at least about 99.8%, at least about 99.9%, at least about 99.91%, at least about 99.92%, at least about 99.93%, at least about 99.94%, at least about 99.95%, at least about
25 99.96%, at least about 99.97%, at least about 99.98% or at least about 99.99% sequence identity with the sequence over its entire length.

The term "treating", "treat" and "therapy", and synonyms thereof as used herein refer to both therapeutic treatment and prophylactic or preventative measures, wherein the object is to prevent or slow down (lessen) a medical
30 condition, which includes but is not limited to diseases (such as SARS-CoV-2 infections), symptoms and disorders. A medical condition also includes a body's response to a disease or disorder, e.g., inflammation. Those in need of such

treatment include those already with a medical condition as well as those prone to getting the medical condition or those in whom a medical condition is to be prevented.

The term "subject" as used herein includes patients and non-patients. The
5 term "patient" refers to individuals suffering or are likely to suffer from a medical condition such as a SARS-CoV-2 infection, while "non-patients" refer to individuals not suffering and are likely to not suffer from the medical condition. "Non-patients" include healthy individuals, non-diseased individuals and/or an individual free from the medical condition. The term "subject" includes humans
10 and animals. Animals may include, but is not limited to, mammals (for example non-human primates, canine, murine and the like), and the like. "Murine" refers to any mammal from the family Muridae, such as mouse, rat, rabbit, and the like.

The term "micro" as used herein is to be interpreted broadly to include dimensions from about 1 micron to about 1000 microns.

15 The term "nano" as used herein is to be interpreted broadly to include dimensions less than about 1000 nm.

The term "particle" as used herein broadly refers to a discrete entity or a discrete body. The particle described herein can include an organic, an inorganic or a biological particle. The particle used described herein may also be a macro-
20 particle that is formed by an aggregate of a plurality of sub-particles or a fragment of a small object. The particle of the present disclosure may be spherical, substantially spherical, or non-spherical, such as irregularly shaped particles or ellipsoidally shaped particles. The term "size" when used to refer to the particle broadly refers to the largest dimension of the particle. For example, when the
25 particle is substantially spherical, the term "size" can refer to the diameter of the particle; or when the particle is substantially non-spherical, the term "size" can refer to the largest length of the particle.

The terms "coupled" or "connected" as used in this description are intended to cover both directly connected or connected through one or more
30 intermediate means, unless otherwise stated.

The term "associated with", used herein when referring to two elements refers to a broad relationship between the two elements. The relationship

includes, but is not limited to a physical, a chemical or a biological relationship. For example, when element A is associated with element B, elements A and B may be directly or indirectly attached to each other or element A may contain element B or vice versa.

5 The term "adjacent" used herein when referring to two elements refers to one element being in close proximity to another element and may be but is not limited to the elements contacting each other or may further include the elements being separated by one or more further elements disposed therebetween.

10 The term "and/or", e.g., "X and/or Y" is understood to mean either "X and Y" or "X or Y" and should be taken to provide explicit support for both meanings or for either meaning.

15 Further, in the description herein, the word "substantially" whenever used is understood to include, but not restricted to, "entirely" or "completely" and the like. In addition, terms such as "comprising", "comprise", and the like whenever used, are intended to be non-restricting descriptive language in that they broadly include elements/components recited after such terms, in addition to other components not explicitly recited. For example, when "comprising" is used, reference to a "one" feature is also intended to be a reference to "at least one" of that feature. Terms such as "consisting", "consist", and the like, may in the appropriate context, be considered as a subset of terms such as "comprising", "comprise", and the like. Therefore, in embodiments disclosed herein using the terms such as "comprising", "comprise", and the like, it will be appreciated that these embodiments provide teaching for corresponding embodiments using terms such as "consisting", "consist", and the like. Further, terms such as "about", "approximately" and the like whenever used, typically means a reasonable variation, for example a variation of +/- 5% of the disclosed value, or a variance of 4% of the disclosed value, or a variance of 3% of the disclosed value, a variance of 2% of the disclosed value or a variance of 1% of the disclosed value.

25 Furthermore, in the description herein, certain values may be disclosed in a range. The values showing the end points of a range are intended to illustrate a preferred range. Whenever a range has been described, it is intended that the range covers and teaches all possible sub-ranges as well as individual numerical

30

values within that range. That is, the end points of a range should not be interpreted as inflexible limitations. For example, a description of a range of 1% to 5% is intended to have specifically disclosed sub-ranges 1% to 2%, 1% to 3%, 1% to 4%, 2% to 3% etc., as well as individually, values within that range such as 1%, 2%, 3%, 4% and 5%. It is to be appreciated that the individual numerical values within the range also include integers, fractions and decimals. Furthermore, whenever a range has been described, it is also intended that the range covers and teaches values of up to 2 additional decimal places or significant figures (where appropriate) from the shown numerical end points. For example, a description of a range of 1% to 5% is intended to have specifically disclosed the ranges 1.00% to 5.00% and also 1.0% to 5.0% and all their intermediate values (such as 1.01%, 1.02% ... 4.98%, 4.99%, 5.00% and 1.1%, 1.2% ... 4.8%, 4.9%, 5.0% etc.) spanning the ranges. The intention of the above specific disclosure is applicable to any depth/breadth of a range.

Additionally, when describing some embodiments, the disclosure may have disclosed a method and/or process as a particular sequence of steps. However, unless otherwise required, it will be appreciated that the method or process should not be limited to the particular sequence of steps disclosed. Other sequences of steps may be possible. The particular order of the steps disclosed herein should not be construed as undue limitations. Unless otherwise required, a method and/or process disclosed herein should not be limited to the steps being carried out in the order written. The sequence of steps may be varied and still remain within the scope of the disclosure.

Furthermore, it will be appreciated that while the present disclosure provides embodiments having one or more of the features/characteristics discussed herein, one or more of these features/characteristics may also be disclaimed in other alternative embodiments and the present disclosure provides support for such disclaimers and these associated alternative embodiments.

DESCRIPTION OF EMBODIMENTS

Exemplary, non-limiting embodiments of viral epitopes are disclosed hereinafter.

Given the unceasing worldwide surge in COVID-19 cases, there is an imperative need to develop highly specific and sensitive serology assays to define exposure to Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2).

Pooled plasma samples from PCR positive COVID-19 patients were used
5 to identify linear B-cell epitopes from a SARS-CoV-2 peptide library of spike (S), envelope (E), membrane (M), and nucleocapsid (N) structural proteins by peptide-based ELISA. Hit epitopes were further validated with 79 COVID-19 patients with different disease severity status, 12 seasonal human CoV, 20 recovered SARS patients and 22 healthy donors.

10 As such, in one aspect of the present disclosure, there is provided an isolated immunogenic polypeptide or epitope comprising at least 15 contiguous amino acids of one or more SARS-CoV-2 protein selected from the group consisting of a spike (S) protein, an envelope (E) protein, a membrane (M) protein, and a nucleocapsid (N) protein.

15 In some examples, the polypeptide comprises and/or consist 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or 15 to 30, or 18 to 24 contiguous amino acids of SARS-CoV-2 protein. In some examples, the SARS-CoV-2 protein sequence includes full-length peptide sequence spanning the spike (S), envelope (E), membrane (M), and nucleocapsid (N). In some
20 examples, the SARS-CoV-2 protein may be obtained from NCBI GenBank accession numbers MN908947. In some examples, the polypeptide as disclosed herein does not include the full-length SARS-CoV-2 protein.

In various examples, the length/size of the polypeptide is no more than about 30, no more than about 29, no more than about 28, no more than about
25 27, no more than about 26, no more than about 25, no more than about 24, no more than about 23, no more than about 22, no more than about 21, no more than about 20, no more than about 19, no more than about 18, no more than about 17, no more than about 16, no more than about 15, no more than about 14, no more than about 13, no more than about 12, no more than about 11, no
30 more than about 10, no more than about 9, no more than about 8, no more than about 7, no more than about 6 or no more than about 5 amino acids long. In

various examples, the length/size of the polypeptide is from about 5 to about 30, from about 10 to about 25 or from about 15 to about 20 amino acids long. In one example, the polypeptide is from about 5 to about 25 amino acids long. In some examples, the length/size of the polypeptide is about 10, about 11, about 12, 5 about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20 or about 21 amino acids long. In various examples, the polypeptide comprises an oligopeptide. In various examples, the polypeptide comprises a short oligopeptide. In various examples, the polypeptide comprises a linear peptide.

In some examples, the polypeptide may be an engineered or isolated 10 SARS-CoV-2 polypeptide. In some example, the polypeptide as disclosed herein is an engineered polypeptide whose amino acid sequence has been designed by man and/or whose existence and/or production require human intervention and/or activity. For example, an engineered SARS-CoV-2 polypeptide may have an amino acid sequence that is intentionally designed to elicit a particular effect 15 and that differs from the amino acid sequences of SARS-CoV-2 polypeptide found in natural coronavirus isolates.

In some examples, the polypeptide comprises a sequence corresponding to the amino acid sequence of the following positions on a SARS-CoV-2 protein (the first amino acid in the spike (S) glycoprotein being annotated as 1): from 20 about 209 to about 226, from about 553 to about 570, from about 769 to about 786, from about 809 to about 826, from about 201 to about 218, from about 209 to about 226, from about 217 to about 234, from about 225 to about 242, from about 233 to about 250, from about 521 to about 538, from about 529 to about 546, from about 537 to about 554, from about 545 to about 562, from about 553 25 to about 570, from about 761 to about 778, from about 769 to about 786, from about 777 to about 794, from about 785 to about 802, from about 793 to about 810, from about 801 to about 818, from about 809 to about 826, from about 817 to about 834, from about 825 to about 842, from about 833 to about 850, from about 641 to about 658, from about 649 to about 666, from about 657 to about 30 674, from about 665 to about 682, from about 673 to about 690, from about 1121 to about 1138, from about 1129 to about 1146, from about 1137 to about 1154,

from about 1145 to about 1162, from about 1153 to about 1170, from about 553 to about 570, from about 769 to 792, from about 809 to about 826, from about 521 to about 538, from about 529 to about 546, from about 537 to about 554, from about 545 to about 562, from about 553 to about 570, from about 801 to about 818, from about 809 to about 826, from about 817 to about 834 from about 825 to about 842, from about 833 to about 850, from about 521 to about 570, from about 801 to about 850. In some examples, the epitope position is as indexed by reference to SARS-CoV-2 amino acid sequence with accession no. MN908947. In various examples, the polypeptide comprises a polypeptide or a subset/variant/fragment thereof sharing at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% or at least about 100% sequence identity with the amino acid sequences or portions thereof, or a polypeptide differing by about one, about two, about three, about four, about five, about six or more amino acids with the amino acid sequences or portions thereof, or portions of the amino acid sequences of the polypeptide or epitope as described herein.

In some examples, the polypeptide comprises a sequence corresponding to the amino acid sequence of the following positions on a SARS-CoV-2 protein (the first amino acid in the nucleocapsid (N) phosphoprotein being annotated as 1): from about 153 to about 170, from about 121 to about 138, from about 129 to about 146, from about 137 to about 154, from about 145 to about 162, from about 153 to about 170, from about 153 to about 174, from about 361 to about 378, from about 369 to about 386, from about 377 to about 394, from about 385 to about 402, from about 393 to about 410, from about 401 to about 418, from about 402 to about 419. In some examples, the epitope position is as indexed by reference to SARS-CoV-2 amino acid sequence with accession no. MN908947. In various examples, the polypeptide comprises a polypeptide or a subset/variant/fragment thereof sharing at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% sequence identity with the amino acid sequences or

portions thereof, or a polypeptide differing by about one, about two, about three, about four, about five, about six or more amino acids with the amino acid sequences or portions thereof, or portions of the amino acid sequences of the polypeptide or epitope as described herein.

5 In some examples, the polypeptide of the present disclosure comprises 18 to 24 contiguous amino acids of spike (S) protein and/or a nucleocapsid (N) protein of SARS-CoV-2. In some examples, the polypeptide of the present disclosure may include, but is not limited to, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, and SEQ ID NO: 60, or a functional variant or fragment thereof; a polypeptide sharing at least 75% sequence identity thereto; or a polypeptide differing by one, two, three, or four amino acids thereto; or portions thereof.

Amongst the polypeptides as disclosed herein, five immunodominant epitopes, S14P5, S20P2, S21P2, N4P5, and S29P4 were identified on the S and N viral proteins. IgG responses to all identified epitopes displayed a strong detection profile, with N4P5 achieving the highest level of specificity (100%) and sensitivity (>96%) against SARS-CoV-2. Furthermore, IgG responses to S14P5, S21P2 and N4P5 were strongly associated with disease severity. Immunodominant linear B cell epitope, S29P4, from the spike protein of SARS-CoV-2 (amino acid residues 1145-1162) was highly recognised by plasma samples from COVID-19 patients, with a specificity and sensitivity profile of

97.6% and 77.8%, respectively. Longitudinal immunoglobulin G (IgG) antibody profiling showed that the epitope remained highly recognised by COVID-19 patients at late recovery timepoint (6 months post-illness onset (pio)). S29P4 lies away from the receptor binding domain (RBD) of the spike glycoprotein, and its use in tandem with current serological assays could allow for greater coverage and detection. This further increases the sensitivity and specificity in detecting viral exposure, which will aid public health institutions during nationwide screenings for COVID-19.

As such, in some examples, the polypeptide comprises an amino acid sequence selected from the group consisting of: SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 55, and SEQ ID NO: 57; or a peptide sharing at least 75% sequence identity thereto; or a peptide differing by one, two, three, or four amino acids thereto; or portions thereof. In some examples, the polypeptide comprises an amino acid sequence selected from the group consisting of: SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 55, and SEQ ID NO: 57; or a peptide sharing at least 75% sequence identity thereto; or a peptide differing by one, two, three, or four amino acids thereto; or portions thereof.

In some examples, the polypeptide shares a sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% sequence identity thereto or portions thereof, or a sequence differing by about one, about two, about three, about four, about five, about six or more amino acids thereto or portions thereof.

In some examples, the polypeptide induces an immune response, optionally the polypeptide induces a B cell immune response.

In some examples, the polypeptide or epitope induces the production of antibodies selected from the group consisting of IgG, IgA, IgM, IgE, and IgD. In some examples, the polypeptide or epitope induces the production of antibodies such as IgG and/or IgM. In some examples, the polypeptide or epitope induces

the production of various IgG isotypes, including, but is not limited to IgG1, IgG2, IgG3, and IgG4.

In some examples, there is provided an isolated immunogenic epitope of at least 12 contiguous amino acids of one or more SARS-CoV-2 protein selected
5 from the group consisting of a spike (S) protein, an envelope (E) protein, a membrane (M) protein, and a nucleocapsid (N) protein.

In some examples, the epitope comprises and/or consist 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or 15 to 50, or 18 to 24
10 contiguous amino acids of SARS-CoV-2 protein. In some examples, the SARS-CoV-2 protein sequence includes full-length peptide sequence spanning the spike (S), envelope (E), membrane (M), and nucleocapsid (N). In some examples, the SARS-CoV-2 protein may be obtained from NCBI GenBank accession numbers MN908947.

In some examples, the epitope as disclosed herein does not include the
15 full-length SARS-CoV-2 protein. In some example, the epitope induces an immune response against SARS-CoV-2. In some example, the epitope induces a B lymphocyte cell immune response against SARS-CoV-2. In some examples, the epitope is a linear B cell epitope of SARS-CoV-2.

In some examples, the polypeptide as described herein may be modified.
20 For example, the peptide may be joined to a wide variety of other oligopeptides or proteins for a variety of purposes. In some examples, the modification may be a post-translational modification, for example by amidation, acetylation, prenylation, carboxylation, glycosylation, pegylation, and the like. In some examples, the modification may also include modifications of glycosylation, for
25 example those made by modifying the glycosylation patterns of a polypeptide during its synthesis and processing or in further processing steps (for example by exposing the polypeptide to enzymes that affect glycosylation, such as mammalian glycosylating or deglycosylating enzymes).

In some examples, the polypeptide as described herein may include or
30 comprise further modifications to improve their resistance to proteolytic

degradation or to optimize solubility properties or to render them more suitable as a therapeutic agent. For example, the polypeptide or epitope as disclosed herein may further include analogs containing residues other than naturally occurring L-amino acids, for example D-amino acids or non-naturally occurring synthetic amino acids. D-amino acids may be substituted for some or all of the amino acid residues.

In some examples, the polypeptide may be modified (and/or synthesized) with a terminal modification that mimics native protein and prevent enzyme degradation. In some examples, the polypeptide may be modified (and/or synthesized) with an amidation of the C-terminus end. In some examples, the polypeptide may be modified (and/or synthesized) with modifications that increases its stability. For example, the polypeptide may be modified by acetylation to increase its stability by preventing degradation. In some examples, the polypeptide may be further modified by biotinylation. For example, the polypeptide may be modified with biotin or biotin-Ahx.

In some examples, the polypeptide or epitope may further comprise a linker. For example, the linker may be a hydrophilic peptide that is added to prevent steric hindrance. In some examples, the linker may be a hydrophilic tetrapeptide added between biotin and the sequence of interest to prevent steric hindrance. In some examples, the linker may comprise the amino acid sequence SGSG (SEQ ID NO: 67). For example, the linker may be a hydrophilic peptide that is added to provide structural flexibility. In some examples, the linker may comprise the amino acid sequence KKKGGGSGGGGSGGGGS (SEQ ID NO: 68).

In some examples, the epitope may be conjugated to other proteins to increase immunogenicity. For example, the peptide may be conjugated with a carrier such as keyhole limpet hemocyanin (KLH) protein or multiple antigenic peptides (MAPS) asymmetric branches. In some examples, the epitopes may be conjugated to nanoparticles. For example, the peptide may be conjugated to gold nanoparticles in a lateral-flow assay.

In some examples, the epitope of 18 contiguous amino acids of spike (S) protein and/or a nucleocapsid (N) protein of SARS-CoV-2.

In some examples, the epitope includes, but is not limited to, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, and SEQ ID NO: 60. In some examples, the epitope as disclosed herein may include a functional variant or fragment thereof of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, and SEQ ID NO: 60.

In some examples, the epitope comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO:

4, SEQ ID NO: 5, SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 55, and SEQ ID NO: 57.

As illustrated in the Experimental Section, the polypeptide of the present disclosure is specific to SARS-CoV-2 and does not recognize seasonal coronavirus or SARS-CoV. In some examples, the polypeptide of the present disclosure is conserved across multiple strains of the SARS-CoV-2. In some examples, the epitope of the present disclosure is not comprised in seasonal coronavirus or SARS-CoV. In some examples, the polypeptide or epitope does not comprise or include an amino acid sequence as set out in S29P4 or SEQ ID NO: 47. In some examples, the seasonal coronavirus may include, but is not limited to, 229E, NL63, OC43, and the like.

In some examples, there is provided nucleic acid molecules encoding the polypeptide or epitope as described herein. The nucleic acid molecules may be used as primers or as probes. That is, in some examples, the nucleic acid molecules may be used as primers in polymerase chain reaction (PCR), as probes for hybridization (including in situ hybridization) and/or as primers for reverse transcription-PCR (RT-PCR). In some examples, the nucleic acids can be DNA or RNA and can be single stranded or double stranded.

Also disclosed are vectors containing the nucleic acid molecules as described herein. In some examples, the vector may be an expression vector containing the nucleic acid molecule as disclosed herein for transformation into a suitable host cell to allow for production of the protein encoded by the nucleic acid constructs.

In some examples, the polypeptides as disclosed herein may be synthesized according to any technique known in the art.

In some examples, the polypeptide or epitope does not include a polypeptide or epitope of seasonal coronavirus or SARS-CoV. In some examples, the polypeptide or epitope of the present disclosure does not include a polypeptide or epitope of one or more SARS-CoV protein selected from the group

consisting of a spike (S) protein, an envelope (E) protein, a membrane (M) protein, and a nucleocapsid (N) protein.

In some examples, the polypeptide is free of (or does not comprise) a sequence corresponding to the amino acid sequence of the following position on a SARS-CoV-2 protein (the first amino acid in the spike (S) glycoprotein being annotated as 1): from about 801 to about 850. In some examples, the polypeptide does not comprise a sequence corresponding to the amino acid sequence of the following positions on a SARS-CoV protein (the first amino acid in the spike (S) glycoprotein being annotated as 1): from about 201 to about 218, from about 537 to about 554, from about 753 to about 770, from about 793 to about 810, from about 761 to about 778, from about 769 to about 786, from about 777 to about 794, from about 785 to about 802, from about 793 to about 810, from about 537 to about 554. In some examples, the epitope position is as indexed by reference to SARS-CoV-2 amino acid sequence with accession no. MN908947. In various examples, the polypeptide comprises a polypeptide sharing at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% sequence identity with the amino acid sequences or portions thereof, or a polypeptide differing by about one, about two, about three, about four, about five, about six or more amino acids with the amino acid sequences or portions thereof, or portions of the amino acid sequences.

In some examples, the polypeptide does not comprise a sequence corresponding to the amino acid sequence of the following positions on a SARS-CoV protein (the first amino acid in the nucleocapsid (N) phosphoprotein being annotated as 1): from about 153 to about 170. In some examples, the epitope position is as indexed by reference to SARS-CoV-2 amino acid sequence with accession no. MN908947. In various examples, the polypeptide comprises a polypeptide sharing at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% sequence identity with the amino acid sequences or portions thereof, or a polypeptide differing by about one, about two, about three, about four, about five,

about six or more amino acids with the amino acid sequences or portions thereof, or portions of the amino acid sequences.

In some examples, the epitope does not include, SEQ ID NO: 6, SEQ ID NO 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 61, SEQ ID
5 NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, or polypeptide that has at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% sequence identity with the amino acid sequences or portions thereof, or a polypeptide differing by about one, about two, about three, about four, about five,
10 about six or more amino acids with the amino acid sequences or portions thereof, or portions of the amino acid sequences.

In another aspect, there is provided a pharmaceutical composition comprising one or more isolated polypeptide as disclosed herein and a pharmaceutically acceptable carrier. In some examples, the pharmaceutical
15 composition may further comprise an adjuvant.

In yet another aspect, there is provided a composition comprising one or more isolated polypeptide as disclosed herein.

In some examples, the pharmaceutical composition as described herein may be an immune modulating composition, immunogenic composition, or
20 vaccine composition. In various examples, there is provided a composition, optionally an immune system modulating composition, further optionally an immune system stimulating/activating composition, further optionally a vaccine composition, comprising a peptide as disclosed herein. The composition may be capable of stimulating, e.g. increasing, amplifying or boosting, an immune
25 response against SARS-CoV-2 in a subject when administered to the subject. In various examples, the composition is capable of increasing an immune response against SARS-CoV-2 in a subject by at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 100%, at
30 least about 2-fold, at least about 3-fold, at least about 4-fold, at least about 5-fold,

at least about 6-fold, at least about 7-fold, at least about 8-fold, at least about 9-fold or at least about 10-fold after administration to the subject. In some examples, an immune response of the subject can be determined by measuring an amount of SARS-CoV-2-antigen binding proteins in a sample from the subject.

- 5 The composition may further comprise an adjuvant. Examples of suitable adjuvants include aluminium hydroxide and aluminium phosphate which may enable a body's immune response to be increased when administered.

In some examples, there is provided a composition, optionally an immune modulating composition, further optionally an immune system stimulating/activating composition, further optionally a vaccine composition comprising a peptide or epitope, including, but is not limited to: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, and SEQ ID NO: 60. In some examples, the epitope as disclosed herein may include a functional variant or fragment thereof of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO:

45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, and SEQ ID NO: 60; or a peptide sharing at least 75% sequence identity thereto; or a peptide differing by one, two, three or four amino acids thereto; or portions thereof.

In some examples, the present invention also provides an immunogenic composition that may include at least one of the following: inactivated virus, virus-like particles, polypeptide as described herein, nucleic acid encoding the polypeptide as described herein or characteristic or biologically active portion thereof, DNA vector that encodes the polypeptide as disclosed herein or characteristic or biologically active portion thereof, and/or expression system, for example, cells expressing one or more SARS-CoV-2 protein to be used as antigens.

In some examples, the polypeptide or epitope as disclosed herein may be used as vaccine assessment tool.

In various examples, there is provided a method of inducing an immune response in a subject, the method comprising administering the composition, optionally an immune modulating composition, further optionally an immune system stimulating/activating composition, further optionally a vaccine composition to the subject.

In yet another aspect, there is provided an agent or an antigen binding molecule capable of specifically binding to the polypeptide or epitope of as described herein.

The agent may be in the form of a binding molecule, peptide (e.g., a polypeptide, an oligopeptide, a protein etc.), a peptidomimetic, an antigen binding protein, an antibody or fragments thereof (e.g., antigen binding fragments), derivatives thereof, combinations thereof, or the like. In various embodiments, a derivative of a molecule is structurally related to the molecule. For example, the derivative may share a common structural feature, fundamental structure and/or

underlying chemical basis with the molecule. A derivative is not limited to one produced or obtained from the molecule although it may be one produced or obtained from the molecule. In some embodiments, the derivative is derivable, at least theoretically, from the molecule through modification of the molecule. In some embodiments, a derivative of a molecule shares or at least retains to a certain extent a function, chemical property, biological property, chemical activity and/or biological activity associated with the molecule. A skilled person will be able to identify, on a case-by-case basis and upon reading of the disclosure, the common structural feature, fundamental structure and/or underlying chemical basis of the molecule that have to be maintained in the derivative to retain the function, chemical property, biological property, chemical activity, and/or biological activity. A skilled person will also be able to identify assays that can prove the retention of the function, chemical property, biological property, chemical activity, and/or biological activity. For example, a binding assay such as ELISA (enzyme-linked immunosorbent assay) may be carried out to determine a binding property of a derivative of a molecule.

In various examples, the agent may be or may comprise an antigen binding protein or antigen binding molecule. An antigen binding protein or molecule may include any protein construct that is capable of binding to SARS-CoV-2 antigen. Examples include, but are not limited to, antibodies and fragments thereof such as antigen binding fragments. Non-limiting examples of antigen binding fragments include one or more fragments or portions of an antibody that retain the ability to specifically bind to an antigen (e.g., SARS-CoV-2), or synthetic modifications of an antibody fragment that retains the desired binding ability to the antigen. In various embodiments, antigen binding fragments include single domain antibodies, further engineered molecules (such as, but is not limited to diabodies, triabodies, tetrabodies, minibodies, and the like), Fab fragments, Fab' fragments, F(ab')₂ fragments, Fd fragments, Fv fragments, single-chain Fv (scFv) molecules, seFv molecules, scFv dimer, BsFv molecules, dsFv molecules, (dsFv)₂ molecules, dsFv-dsFv' molecules, Fv fragments, dAb fragments, bispecific antibodies, ds diabodies, nanobodies, domain antibodies, bivalent domain antibodies, and minimal recognition units consisting of the amino acid

residues that mimic the hypervariable region of an antibody (e.g. an isolated complementarity determining region (CDR)). In various examples, antigen binding molecule is one selected from the group consisting of IgG, IgM, IgE, IgA, IgD, F(ab) fragments, and fragments thereof. The antigen binding molecule may
5 be an immunoglobulin and may be one or more of IgA, IgM, IgG, IgG1, IgG2, IgG3 and IgG4, human IgA, human IgM, human IgG, human IgG1, human IgG2, human IgG3 and human IgG4.

In the Experimental Section, five immunodominant peptides are presented: S14P5, S20P2, S21P2, N4P5, and S29P4, located on the S and N
10 proteins of SARS-CoV-2. The present disclosure also provides the best peptide combination with the highest specificity and sensitivity for determination of SARS-CoV-2 exposure. The magnitude of the humoral responses against these epitopes shows a strong correlation with disease severity.

As such, in yet another aspect, there is provided a method of detecting the
15 presence of SARS-CoV-2 virus in a sample comprising contacting the sample to one or more polypeptide as described herein and determining the binding of a component of the sample to the polypeptide, wherein the binding of a component of the sample to the polypeptide is indicative of the presence of SARS-CoV-2.

In some examples, the component of the sample may be a binding
20 molecule, such as an antigen binding protein or molecule as described herein.

In yet another aspect, a method of predicting the severity of a SARS-CoV-2 infection comprising contacting the sample to one or more polypeptide selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 50, SEQ ID NO: 55, and SEQ ID NO: 57 and determining the binding of
25 a component of the sample to the polypeptide, wherein the binding of a component of the sample to the polypeptide is indicative of an increased risk of a severe SARS-CoV-2 infection.

In the present disclosure, the outcome of a SARS-CoV-2 infection may be categorized into three scores as follows:

- 30 i. mild outcome or score 0, which subject presents fever and flu-like symptoms with no pneumonia,

- ii. moderate outcome or score 1, which subject presents pneumonia with no hypoxia, or
- iii. severe outcome or score 2, which subject presents pneumonia with hypoxia.

5 In some examples, the method may further comprise determining the levels of one or more blood parameters such as, but not limited to, C-reactive protein (CRP) and lactate dehydrogenase (LDH) levels, and lymphocyte count.

In some examples, the sample comprises a biological sample, optionally the sample is a fluid biological sample.

10 In some examples, the method of detecting the presence of SARS-CoV-2 virus in a sample may be performed using standard immunological assays to assess whether antigen binding proteins (e.g. antibodies) that are capable of recognizing or binding to the peptide are present in the sample. Examples of immunoassay include radioimmunoassay (RIA), chemiluminescence- and
15 fluorescence-immunoassays, enzyme-linked immunosorbent assay (ELISA), enzyme-linked immunosorbent spot (ELISPOT), Luminex-based bead arrays, protein microarray assays, immunochromatography (ICG)-based assay and rapid test formats such as lateral-flow assay or immunochromatographic strip tests. The assay can be competitive and non-competitive sandwich assays.

20 In some examples, the sample is obtained from a subject. In some examples, the subject may be a non-human animal and human. In some examples, the subject may be a human or a non-human mammal such as, but is not limited to, bats, rodents, non-human primates, equines, and the like.

In various examples, the sample comprises a biological sample. In various
25 examples, the biological sample comprises a fluid biological sample or a liquid biological sample. In some examples, the biological sample may be a nasal swab. The fluid biological sample or liquid biological sample may be blood, serum, plasma, sputum, lavage fluid, cerebrospinal fluid, urine, semen, sweat, tears, saliva, and the like. In some examples, the fluid biological sample or liquid
30 biological sample comprises whole blood, blood serum, blood plasma or

processed fractions thereof. In some examples, the fluid biological sample comprises blood serum or blood plasma. In some examples, the fluid biological sample comprises antigen binding proteins such as antibodies.

In various examples, the sample comprises a sample that is collected from
5 a subject from about 2 days to about 2 years post illness onset (pio). In some
examples, the sample comprises a sample that is collected from a subject from
about 2 days to about 60 days pio, or from 2 days up to 6 months pio, or from 2
days up to 2 years pio, or about 6 months pio, or about 1 year pio, or about 1.5
10 years pio, or about 2 years pio, or more. In some examples, the sample is
collected from a subject when subject shows symptoms associated with SARS-
CoV-2 infection, such as fever, chills, cough, shortness of breath or difficulty in
breathing, fatigue, muscle or body ache, headache, new loss of taste or smell,
sore throat, congestion or runny nose, nausea or vomiting, and diarrhea. In some
examples, the sample is collected from a subject when subject is or has become
15 asymptomatic for SARS-CoV-2 infection. In some examples, the sample may be
collected from a subject who has received one or more doses of vaccine against
SARS-CoV-2. In some examples, the sample may be collected from a subject
from about 2 days to about 60 days post-vaccination, or from 2 days up to 6
months post-vaccination, or from 2 days up to 2 years post-vaccination, or about
20 6 months post-vaccination, or about 1-year post-vaccination, or about 1.5 years
post-vaccination, or about 2 years post-vaccination, or more.

In various examples, the method has high sensitivity and/or specificity. In
various examples, the method has a sensitivity of at least about 50%, at least
about 55%, at least about 60%, at least about 65%, at least about 70%, at least
25 about 75%, at least about 80%, at least about 81%, at least about 82%, at least
about 83%, at least about 84%, at least about 85%, at least about 86%, at least
about 87%, at least about 88%, at least about 89%, at least about 90%, at least
about 91%, at least about 92%, at least about 93%, at least about 94%, at least
about 95%, at least about 96%, at least about 97%, at least about 98%, at least
30 about 99% or at least about 100%.

In various examples, the method has a specificity of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 81%, at least about 82%, at least about 83%, at least about 84%, at least about 85%, at least about 86%,
5 at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% or at least about 100%.

10 In various examples, the method comprises an in vitro or an ex vivo method.

In some examples, the method may identify/distinguish a current or previous SARS-CoV-2 infection from other infections, for example any previous SARS-CoV infection, other coronaviral or seasonal viral infection.

15 In some examples, there is provided an isolated nucleic acid encoding the polypeptide as described herein. In some examples, there is provided a vector comprising the nucleic acid encoding the polypeptide as described herein. In some examples, there is provided a host cell transformed or transfected with the vector as described herein.

20 In some examples, there is provided a method of producing the polypeptide as described herein, wherein said method comprises culturing a host cell as described herein under conditions permitting expression of the polypeptide.

25 In addition, these five epitopes could be further developed into point-of-care tests (POCTs) for detection of different immunoglobulins (such as, but not limited to, IgG, IgA and IgM) and IgG isotypes. As such, in yet another aspect, there is provided a kit comprising one or more polypeptide as disclosed herein and/or the antigen binding molecule as disclosed herein.

In some examples, the kit may further comprise one or more of the following: a substrate coated with a capture agent for anti-SARS-CoV-2 antigen

binding molecule, and a detection agent for detecting the presence of captured anti-SARS-CoV-2 antigen binding molecule, wherein the capture agent and the detection agent comprise a SARS-CoV-2 antigen and/or anti-immunoglobulin.

The capture and detection agent may be independently selected from a
5 purified SARS-CoV-2 antigen, antibody of anti- SARS-CoV-2 or an anti-immunoglobulin. The anti-immunoglobulin may be one or more of anti-IgA, anti-IgM, anti-IgG, anti-IgG1, anti-IgG2, anti-IgG3 and anti-IgG4, anti-human IgA, anti-human IgM, anti-human IgG, anti-human IgG1, anti-human IgG2, anti-human IgG3 and anti-human IgG4. In some examples, the capture agent comprises
10 SARS-CoV-2 antigen and the detection agent comprises a labelled anti-immunoglobulin. In some examples, the capture agent comprises anti-immunoglobulin and the detection agent comprises labelled SARS-CoV-2 antigen. In some examples, the capture agent and the detection agent comprise a SARS-CoV-2 antigen and/or an anti-immunoglobulin. In some examples, the
15 capture agent and the detection agent are selected from the group consisting of:
(i) a SARS-CoV-2 antigen and one or more of a labelled anti-immunoglobulin and
(ii) an anti-immunoglobulin and a labelled SARS-CoV-2 antigen.

In some examples, the kit further comprises one or more of the following:
a plate coated with a capture agent for anti-SARS-CoV-2, and a detection agent
20 for detecting the presence of captured anti- SARS-CoV-2, wherein the capture agent and the detection agent comprise a SARS-CoV-2 antigen and/or an anti-immunoglobulin.

The kit may be a diagnostic kit or a prognostic kit.

In yet another aspect, there is provided a method of treating or preventing
25 SARS-CoV-2 in a subject in need thereof comprising administering into the subject a polypeptide as described herein.

In yet another aspect, there is provided a polypeptide as described herein for use as a medicament. Also provided is the use of the polypeptide as described herein in the manufacture of a medicament for use in treating SARS-CoV-2 in a
30 subject in need thereof.

In various examples, there is provided a method, a kit, a peptide or a composition as described herein.

DETAILED DESCRIPTION OF FIGURES

5

Example embodiments of the disclosure will be better understood and readily apparent to one of ordinary skill in the art from the following discussions and if applicable, in conjunction with the figures. It should be appreciated that other suitable modifications may be made without deviating from the scope of the invention. Example embodiments are not necessarily mutually exclusive as some may be combined with one or more embodiments to form new exemplary embodiments. The example embodiments should not be construed as limiting the scope of the disclosure.

FIG. 1 Identification of IgG-specific linear B-cell epitopes on spike (S), envelope (E), membrane (M), and nucleocapsid (N) SARS-CoV-2 proteins with COVID-19 patient plasma. (A-C) Plasma samples of COVID-19 patients collected at timepoint of median 14 days post-illness onset (pio, n=18), recovered SARS individuals (n=20) and healthy donors (n=22) were pooled, respectively. **(A)** Pooled plasma samples (1:1000 dilution) were tested in a peptide-based ELISA of peptide pools covering SARS-CoV-2 S, E, M, and N proteins for IgG. Each pool consists of 5-8 peptides, with an overlapping sequence of 10 amino acids. OD results are presented in a heatmap, with white and grey colours denoting low and high OD values, respectively. The cut-off for positive pools was derived from mean + 3SD of pooled healthy donors. **(B)** Individual peptides of the positive peptide pools in **(A)** were screened with pooled plasma samples (1:1000 dilution) to determine the IgG hit peptides. OD results are presented in a heatmap, with white and grey colours denoting low and high OD values, respectively. **(C)** Peptide-binding response of pooled plasma samples (1:1000 dilution) against the corresponding hit peptides on SARS-CoV, with bar graphs presented as mean $\pm$ SEM, and dotted line denoting mean + 3SD of pooled healthy donors. All data are of two independent experiments. **(D-E)** Diagrams

showing the localisation of SARS-CoV-2 specific IgG peptides S6P2, S14P5, S20P2, S21P2, and N4P5 on **(D)** SARS-CoV-2 S (PDB: 6VSB), and **(E)** N (PDB: 6YVO) protein.

FIG. 2 Specificity and sensitivity of putative IgG epitopes to detect anti-SARS-CoV-2 antibodies from COVID-19 patients. **(A)** Plasma samples from COVID-19 patients collected at timepoints of median 10 days post-illness onset (pio, n=59) and median 23 days pio (n=66), sera from seasonal hCoV-infected patients (n=12) either pre- or post-infection, and plasma samples from recovered SARS patients (n=20) and healthy donors (n=22) were screened at 1:1000 dilution against five IgG-specific putative epitopes: S6P2, S14P5, S20P2, S21P2, and N4P5. Data are shown as mean $\pm$ SD of two independent experiments, with dotted lines indicating mean + 3SD of healthy donors. **(B)** Percentage specificity and sensitivity of putative epitopes to detect SARS-CoV-2-specific antibodies in plasma samples from COVID-19 patients during the timepoints of median 10 days pio (left panel) and median 23 days pio (right panel). Specificity and sensitivity values were derived from a threshold determined by the maximisation of the Youden's J statistic with a constraint that the threshold value has to be positive. **(C)** Percentage recognition of epitopes during the timepoints of median 10 days pio (black bars) and median 23 days pio (white bars).

FIG. 3 Association of anti-peptide IgG response of COVID-19 patients with clinical severity during SARS-CoV-2 infection. Antibody profiles of COVID-19 patients at timepoint of median 23 days post-illness onset (pio, n=66) against significant IgG peptides under different clinical measures of **(A)** disease severity, **(B)** pneumonia, and **(C)** intensive care unit (ICU) admission and oxygen supply requirement. Samples in **(A)** were classified based on severity of mild (score 0; n=28), moderate (score 1; n=22), and severe (score 2; n=16). Data are presented as mean $\pm$ SD of two independent experiments. Negative and positive observations in **(B-C)** are denoted as - and +, respectively. Statistical analysis was carried out using Kruskal-Wallis tests, followed by post-hoc Dunn's multiple comparison tests. (** P <0.01, *** P <0.001).

FIG. 4 An example of a strip plot showing the peptide OD value on the Y-axis, and the sample groups on the X-axis. Each dot represents a unique biological sample being either a COVID-19 sample (positive) or a non-COVID-19 sample (negative). The true positive (TP), true negative (TN), false positive (FP) and false negative (FN) samples are highlighted using different shaded backgrounds. The OD threshold is depicted as a dotted line.

FIG. 5 An example of a receiver operating characteristic (ROC) curve showing the best threshold (value of 0.291) with an achieved sensitivity of 0.763 and specificity of 0.805. The impact of maximising the Youden J statistics is shown in the figure as a double-head arrow, which clearly indicates the threshold selected as the maximum perpendicular distance from the diagonal line.

FIG. 6 Identification of IgM-specific linear B-cell epitopes on spike (S), envelope (E), membrane (M), and nucleocapsid (N) SARS-CoV-2 proteins with COVID-19 patient plasma. (A-C) Plasma samples of COVID-19 patients collected at timepoints of median 5 days post-illness onset (pio, n=6) and median 14 days pio (n=18), recovered SARS individuals (n=20) and healthy donors (n=22) were pooled, respectively. **(A)** Pooled plasma samples (1:1000 dilution) were tested in a peptide-based ELISA of peptide pools covering SARS-CoV-2 S, E, M, and N proteins for IgM. Each pool consists of 5-8 peptides, with an overlapping sequence of 10 amino acids. The cut-off for positive pools was derived from mean + 3SD of pooled healthy donors. **(B)** Individual peptides of the positive peptide pools in **(A)** were screened with pooled plasma samples (1:1000 dilution) to determine the IgM hit peptides. OD results of **(A-B)** are presented in a heatmap, with white and grey colours denoting low and high OD values, respectively. **(C)** Peptide-binding response of pooled plasma samples (1:1000 dilution) against the corresponding hit peptides on SARS-CoV, with bar graphs presented as mean $\pm$ SEM, and dotted line denoting mean + 3SD of pooled healthy donors. **(D)** Localisation of IgM peptide S20P4 on SARS-CoV-2 S protein (PDB: 6VSB). **(E)** Plasma samples from COVID-19 patients at timepoints median 10 days pio (n=59) and median 23 days pio (n=66), sera from seasonal hCoV-infected patients (n=12) either pre- or post-infection, and plasma samples from recovered SARS patients (n=20) and healthy donors (n=22) were screened at

1:1000 dilution against IgM peptides S20P4 and N10P1. Data are shown as mean $\pm$ SD with dotted lines indicating mean + 3SD of healthy donors. All data are of two independent experiments.

FIG. 7 Anti-SARS-CoV-2 antibody response of COVID-19 patients against IgG epitopes during SARS-CoV-2 infection. (A) Antibody profiling of individual plasma samples (1:1000 dilution) from COVID-19 patients at timepoints of median 10 days post-illness onset (pio, n=59) and median 23 days pio (n=66) against IgG peptides S14P5, S20P2, S21P2 and N4P5. Samples were classified based on severity of mild (score 0; n=23 for median 10 days, n=28 for median 23 days pio), moderate (score 1; n=22 for both median 10 and 23 days pio), and severe (score 2; n=14 for median 10 days, n=16 for median 23 days pio). Data are presented as mean $\pm$ SD of two independent experiments. (B-D) Antibody profiles of COVID-19 patients at timepoint of median 23 days pio (n=66) against significant IgG peptides under different laboratory parameters of (B) C-reactive protein (CRP) and (C) lactate dehydrogenase (LDH) levels, and (D) lymphocyte count. Spearman correlation with rho (ρ) values and P_{adj} values are presented. (* P <0.05, ** P <0.01, *** P <0.001).

FIG. 8 Preliminary mapping of linear B cell pooled epitopes within the SARS-CoV-2 spike protein. Pooled plasma samples from COVID-19 patients at 1, 3 and 6 months post-illness onset (pio) were subjected to peptide-based ELISA (in duplicates) for IgG detection at 1:100 plasma dilution against 31 pooled peptides of the SARS-CoV-2 spike protein. Heatmaps showing the IgG levels from COVID-19 patients of (A) low (<50% IgG recognition) and (B) high (>50% IgG recognition) binders (n=8 per group, except low binders at 1 month pio, n=7). (C) Heatmap showing the IgG levels from pooled plasma samples of recovered SARS subjects (n=20) and healthy donors (old, n=14; young, n=8). Data are presented as mean of two independent experiments.

FIG. 9 Determination of the specific linear B cell epitope on pool 17 of SARS-CoV-2 spike protein. Plasma samples from COVID-19 patients at 1, 3 and 6 months post-illness onset (pio) were subjected to peptide-based ELISA (in duplicates) for IgG detection at 1:100 plasma dilution against individual peptides of pool 17 of SARS-CoV-2 spike protein. COVID-19 patients were stratified into

low (<50% IgG recognition) and high (>50% IgG recognition) binders (n=8 per group, except low binders at 1 month pio, n=7). Recovered SARS (n=20), old (n=14) and young healthy donors (n=8) were screened in parallel. Data showing the IgG levels of **(A)** pooled plasma samples against five peptides within pool 17 and **(B)** individual plasma samples against selected peptides of pool 17. Data are presented as mean $\pm$ SD of two independent experiments. Dotted line indicates mean + 3SD of healthy controls.

FIG. 10 Determination of the specific linear B cell epitope on pool 29 of SARS-CoV-2 spike protein. Plasma samples from COVID-19 patients at 1, 3 and 6 months post-illness onset (pio) were subjected to peptide-based ELISA (in duplicates) for IgG detection at 1:100 plasma dilution against individual peptides of pool 29 of SARS-CoV-2 spike protein. COVID-19 patients were stratified into low (<50% IgG recognition) and high (>50% IgG recognition) binders (n=8 per group, except low binders at 1 month pio, n=7). Recovered SARS (n=20), old (n=14) and young healthy donors (n=8) were screened in parallel. Data showing the IgG levels of **(A)** pooled plasma samples against five peptides within pool 29 and **(B)** individual plasma samples against peptide 4 of pool S29. Data are presented as mean $\pm$ SD of two independent experiments. Dotted line indicates mean + 3SD of healthy controls. **(C-D)** Localisation of peptide 4 of pool 29 (S29P4), receptor binding domain (RBD) and fusion peptide on SARS-CoV-2 spike protein in **(C)** open (PDB: 6VSB) and **(D)** closed (PDB: 6X79) conformations.

FIG. 11 Detection of IgG binding antibodies in COVID-19 patients against S29P4. **(A)** Plasma samples from COVID-19 patients at 1 (n=106), 3 (n=101) and 6 (n=103) month post-illness onset (pio) were subjected to peptide-based ELISA against S29P4 at 1:100 dilution in duplicates. Plasma samples from recovered SARS patients (n=20) and healthy controls (n=22) were also tested in parallel. Data are shown as mean $\pm$ SD of two independent experiments. Dotted line represents mean + 3SD of healthy controls. Statistical tests were performed using Kruskal-Wallis with Dunn's multiple comparison post-hoc test (**P<0.01, ****P<0.0001). **(B)** Receiver operating characteristic (ROC) profile of S29P4 at 1 month pio. AUC, area under curve.

FIG. 12 Comparison of IgG recognition between S14P5, S20P2, S21P2, N4P5 and S29P4 at recovery timepoints. (A) COVID-19 patient plasma samples at 1 (n=106), 3 (n=101) and 6 (n=103) months post-illness onset (pio) were subjected to peptide-based ELISA against previously characterised epitopes, S14P5, S20P2 (extended sequence), S21P2 and N4P5 (extended sequence), and S29P4 at 1:100 dilution. A cut-off is set at mean + 3SD of healthy controls for each individual epitope, with samples above cut-off (positive recognition) in squared symbols, and samples below cut-off (negative recognition) in circled symbols. Data are shown as mean $\pm$ SD of two independent experiments. **(B)** Percentage recognition calculated for epitopes at 1, 3 and 6 months pio. This is represented as the percentage of samples with OD450nm values above cut-off (set at mean + 3SD of healthy controls).

FIG. 13 Preliminary mapping of linear B cell pooled epitopes within SARS-CoV-2 and SARS-CoV spike protein. Serum samples of recovered COVID-19 patients (n=6) at 1:1000 dilution were subjected to peptide-based ELISA for IgG using pooled peptides of the spike protein of **(A)** SARS-CoV-2 (pools 1-31) and **(B)** SARS-CoV (pools 32-62) in duplicates. Recovered SARS subjects (n=5) and pooled healthy donors (n=13) were also tested in parallel. Data are presented as mean $\pm$ SD of OD450 nm after subtraction with pooled healthy donors (n=13).

FIG. 14 Determination of SARS-CoV-2 specific linear B cell epitope on spike protein. (A-B) Serum samples of recovered COVID-19 patients (n=6) at 1:1000 dilution were subjected to peptide-based ELISA (in duplicates) for IgG **(A)** using individual peptides of spike SARS-CoV-2 pooled peptide 14. The z-score values of patient were calculated using the formula [value – average (OD value of patient) / standard deviation (OD value of patient)]. Data are of two independent experiments and presented as mean. **(B)** Comparison of binding response of COVID-19 recovered patients to SARS-CoV-2 spike peptide 14-5 and the corresponding peptide region on SARS-CoV peptide 45-3. Statistical analysis was carried out with paired parametric two-tailed t-test (*p<0.05). **(C)** Localization of SARS-CoV-2 specific epitope (peptide 14-5) and ACE2 binding region on spike protein (PDB: 6VSB).

FIG. 15 Determination of pan-coronavirus linear B cell epitope on spike protein. (A) Serum samples of recovered COVID-19 patients (n=6) at 1:1000 dilution were subjected to peptide-based ELISA (in duplicates) for IgG using individual peptides of spike SARS-CoV-2 pooled peptide 21. The z-score values of patient were calculated using the formula $[\text{value} - \text{average (OD value of patient)}] / \text{standard deviation (OD value of patient)}$. Data are of two independent experiments and presented as mean. **(B)** Comparison of binding response of COVID-19 recovered patients to SARS-CoV-2 spike peptide 21-2 and the corresponding peptide region on SARS-CoV peptide 51-5. **(C)** Localization of pan-coronavirus epitope (peptide 21-2) and fusion peptide on spike protein (PDB: 6VSB). Underlined sequences indicate the overlapping region between peptide 21-2 and the fusion peptide.

FIG. 16 Evaluation of peptide derivatives of identified immunodominant linear B cell epitopes of SARS-CoV-2. Pooled plasma samples collected from COVID-19 patients at the convalescent (mild, n=73; moderate, n=52; severe, n=56; median 14 days post-illness (pio)) and recovered (mild, n=44; moderate, n=34; severe, n=19; median 23 days pio) phases, recovered SARS patients (n=20), old (n=8) and young (n=14) healthy donors were subjected to peptide-based ELISA against **(A)** identified immunodominant linear B cell epitopes of SARS-CoV-2, namely S14P5, S20P2, S21P2 and N4P5. Data shown are from two independent experiments and represented as mean $\pm$ SD. **(B)** Sequences of peptide derivatives used in **(A)**.

EXPERIMENTAL SECTION

25

Materials and Methods

Ethical approvals

The study design and protocols for COVID-19, recovered SARS and seasonal hCoV patient cohorts were evaluated by National Healthcare Group (NHG) Domain Specific Review Board (DSRB) and approved under study numbers 2012/00917, 2020/00091 and 2020/00076, respectively. Healthy donor

30

samples were collected under study numbers 2017/2806 and NUS IRB 04-140. Residual serum sample from National Healthy Survey (NHS) collected in 2010 under study number 006/2010 was used as a positive control in peptide-based ELISA (13). Written informed consent was obtained from participants in accordance with the Declaration of Helsinki for Human Research.

Patients and sample collection

COVID-19 patients

Eighty-one patients who tested PCR-positive for SARS-CoV-2 in nasopharyngeal swab samples were recruited into the study from January to March 2020 (12). Demographic data, disease onset date, clinical score and SARS-CoV-2 RT-PCR results during the hospitalisation period were retrieved from patient clinical records (Table 1). Patients were classified into three groups based on clinical severity: mild (no pneumonia; clinical score 0), moderate (pneumonia without hypoxia; clinical score 1), and severe (pneumonia with hypoxia; clinical score 2). Whole blood of patients was collected into BD Vacutainer® CPT™ tubes (BD Biosciences), and centrifuged at 1700 g for 20 min to obtain plasma fractions. Plasma samples were categorised according to three timepoints: median 5 days pio, median 10-14 days pio, and median 23 days pio.

Table 1: Demographic and clinical data of COVID-19 patients

	Patients (n=81)
Demographics	
Age, years	45 (13)
Sex	
Male	48 (59.3%)
Female	33 (40.7%)
Ethnicity	
Chinese	68 (84.0%)
Others	13 (16.0%)
Comorbidities	28 (34.6%)
Diabetes	7 (8.6%)
Hypertension	15 (18.5%)
Others	11 (13.6%)
Vital signs at admission	
Temperature, °C	37.7 (0.9)
Heart rate, beats/minute	91.4 (16.6)

Respiratory rate, per minute	18.4 (1.9)
Diastolic blood pressure, mmHg	97.5 (2.4)
Systolic blood pressure, mmHg	132.2 (18.5)
Oxygen saturation, %	77.8 (15.2)
Laboratory findings	
Haemoglobin, g/dL	13.8 (1.6)
Haematocrit, %	40.8 (4.6)
Platelets, x 10 ⁹ /L	194.8 (69.8)
White blood cells, x 10 ⁹ /L	5.3 (3.0)
Lymphocytes, x 10 ⁹ /L	1.2 (0.6)
Neutrophils, x 10 ⁹ /L	4.3 (7.7)
Monocytes, x 10 ⁹ /L	0.6 (1.1)
C-reactive protein (CRP), mg/L	37.4 (55.7)
Creatinine, µmol/L	75.0 (45.3)
Lactate dehydrogenase (LDH), U/L	514.3 (298.4)
Alanine aminotransferase (ALT), U/L	34.6 (28.1)
Clinical outcome (severity group)	
No pneumonia (0; mild)	34 (42.0%)
Pneumonia, without hypoxia (1; moderate)	28 (34.5%)
Pneumonia, with hypoxia (2; severe)	19 (23.5%)
Data represented as Mean (S·D) or n (%). COVID-19: Coronavirus Disease-19.	

Recovered SARS and seasonal hCoV patients

A total of 20 individuals who survived SARS-CoV infections in the 2003 outbreak were enrolled. Plasma fractions from recovered SARS individuals were isolated as described above. Archived paired samples from hCoV patients collected between 2012-2013 were also used in this study. Pre- and post-infected serum samples from seven alpha-CoV (229E/NL63) and six beta-CoV positive (OC43) patients were obtained.

10

Design of linear peptide libraries

SARS-CoV-2 and SARS-CoV full-length peptide sequences spanning the spike (S), envelope (E), membrane (M) and nucleocapsid (N) were obtained from NCBI GenBank accession numbers MN908947 and NC_004718, respectively. Biotinylated linear peptides of 18-mer overlapping sequences of 10 residues were synthesised (Mimotopes) and reconstituted in DMSO (Sigma-Aldrich). An initial

15

screen was conducted using pooled peptide sets of five to eight peptides per pool.

Plasma inactivation and peptide-based ELISA

5 Plasma fractions were treated with Triton X-100 (ThermoFisher Scientific) to a final concentration of 1% for 2 h at room temperature (RT) for virus inactivation (14). Epitope screening was performed using a peptide-based ELISA as previously described (15-18). Briefly, Maxisorp flat-bottom 96-well plates (ThermoFisher Scientific) were coated overnight at 4°C with 1:2000 dilution of
10 NeutrAvidin protein (1 mg/ml) (ThermoFisher Scientific) in PBS. Plates were blocked with a 0.01% Polyvinyl Alcohol (PVA) (Sigma-Aldrich) solution in 0.1% PBST (blocking buffer) before addition of pooled or single biotinylated peptides (1:2000 dilution in 0.1% PBST), and inactivated plasma samples (1:1000 dilution in 0.1% PBST). Goat anti-human IgM-HRP or IgG-HRP (Jackson
15 ImmunoResearch) diluted in blocking buffer were used for detection of peptide-bound antibodies. For development, tetramethylbenzidine (TMB) substrate (Sigma-Aldrich) was added to the plates and the reaction was stopped using 0.16 M sulfuric acid (Merck). Absorbance measurements were done using two wavelengths (450 nm and 690 nm) on an Infinite M200 plate reader (Tecan,
20 firmware V_2.02_11/06). In all steps, plates were incubated at RT for 1 h on a rotating shaker and washed twice with 0.1% PBST in between steps.

Sequence alignment and computational modelling

Sequence alignment was conducted for each structural protein of SARS-CoV-2 and SARS-CoV S, E, M and N using Clustal Omega (EMBL-EBI, version
25 1.2.4). Structural data of immunodominant regions on SARS-CoV-2 S and N proteins were retrieved from PDB 6VSB and 6YVO and were modelled in PyMol (Schrodinger, version 2.2.0).

30 Data and statistical analyses

Optical density (OD) values of samples were normalised to a positive control to account for plate-to-plate variations, and background signals were

subtracted. In the first screening phase of the peptide pools, a cut-off value of mean + 3SD of healthy control samples was used to single out for individual peptides detected by COVID-19 patients. For subsequent analyses with individual peptides, receiver operating characteristic (ROC) curves for peptides to differentiate between SARS-CoV-2 infections and others was performed, and the areas under the curve (AUCs) were calculated for each peptide (supplementary methods). The best threshold for each ROC curve was determined as the maximum of the Youden's J statistic (19) with a constraint that the threshold has to be a positive value (supplementary methods). For peptide combination analyses, a logistic regression model was used to differentiate between SARS-CoV-2 infection from others to determine the AUCs. Data analyses were done using GraphPad Prism (GraphPad Software, version 8.0.0). Kruskal-Wallis tests and post hoc tests using Dunn's multiple comparison tests were used to identify significant differences. Correlation analysis was carried out using Spearman's rank correlation. *P*-values less than 0.05 are considered significant.

For peptide-based ELISA assays, OD values were obtained from COVID-19 samples, with non-COVID-19 samples serving as the negative controls. An OD threshold can be used to determine the assay positivity, which will allow for the classification of samples into true positives (TP), true negatives (TN), false positive (FP) and false negative (FN), as depicted in FIG 4.

The best threshold for each peptide assay was determined by maximising the Youden J statistics (19), from which, both specificity and sensitivity could be determined. Receiver Operating Characteristics (ROC) curves were used to visualise the relationship between the sensitivity and specificity for all possible OD thresholds as shown in FIG 5. The best threshold selected is shown in the figure, but the ROC allows readers to tune the assay for more specificity or sensitivity as required by the clinical usage situation.

30

Results

Identification of immunodominant linear B-cell epitopes on SARS-CoV-2 S and N proteome

In order to identify potential linear B-cell epitopes, pooled plasma samples from 18 COVID-19 patients of the Singapore outbreak (12) (Table 1) collected during the convalescent phase (median 14 days post-illness onset) were screened by peptide-based ELISA for IgG reactivity against a pooled peptide library encompassing the entire SARS-CoV-2 structural proteome (FIG. 1A). Freshly drawn plasma samples of 20 recovered SARS individuals and 22 healthy donors were also assessed in parallel (FIG. 1A). While there was no recognition on the E and M peptide pools by the COVID-19 patient samples (FIG. 1A), several hits were identified on regions in the S and N proteins (5, 8, 20, 21). Positive pools S6, S14, S20, S21 from S glycoprotein, and N4 from N protein, were chosen for further characterisation of the specific peptide recognised (FIG. 1A).

Within each pool, distinct peptides highly recognised by IgGs present in the plasma samples of COVID-19 patients were identified: peptide S6P2 from pool S6, peptide S20P2 from pool S20, and peptide N4P5 from pool N4 (FIG. 1B, Table 2 and Table 5). Two reported peptide epitopes, S14P5 and S21P2, were the dominant hits from their respective pools (FIG. 1B, Table 2 and Table 5) (18). It is also noteworthy that IgG levels of recovered SARS and healthy individuals in all of the assays were very low or negligible, highlighting that COVID-19 patients recognised these peptides with high specificity (FIG. 1A-B).

Given the high level of homology between SARS-CoV-2 and the closely-related SARS-CoV (8, 22), pooled COVID-19 patient plasma samples were also screened against the corresponding hit regions on SARS-CoV S and N proteins to assess for potential cross-reactivity (FIG. 1C and Table 2). Interestingly, while COVID-19 patient plasma samples were able to recognise SARS-CoV versions of peptides S20P2 and N4P5, there was no binding to S6P2, alluding to the identification of a potential SARS-CoV-2 specific epitope (FIG. 1C). The localisation of all the putative epitopes on the respective viral proteins is shown in FIG. 1D-E. All epitopes are located on the surface of the viral protein, except for S20P2, which is partially buried in the S protein 3D structure (FIG. 1D).

Table 2: Peptide sequences of putative immunodominant SARS-CoV-2 IgG linear B-cell epitopes

	SARS-CoV-2				SARS-CoV			Identity with SARS-CoV-2 (%)
Peptide	Starting aa	Ending aa	Polypeptide sequence	Starting aa	Ending aa	Polypeptide sequence		
Spike (S) glycoprotein								
S6 P2	209	226	PINLVRDLPPQGFSALEPL (SEQ ID NO: 1)	201	218	QPIDVVRDLPSGFNTLKP (SEQ ID NO: 6)	64.71	
S14 P5	553	570	TESNKKFLPFQQFGRDIA (SEQ ID NO: 2)	537	554	VLTSSSKRFQPFQQFGRD (SEQ ID NO: 7)	75.00	
S20 P2	769	786	GIAVEQDKNTQEVFAQVK (SEQ ID NO: 3)	753	770	AAEQDRNTREVFAQVKQM (SEQ ID NO: 8)	81.25	
S21 P2	809	826	PSKPSKRSFIEDLLFNKV (SEQ ID NO: 4)	793	810	KPTKRSFIEDLLFNKVTL (SEQ ID NO: 9)	93.75	
Nucleocapsid (N) phosphoprotein								
N4 P5	153	170	NNAAIVLQLPQGTTLPKG (SEQ ID NO: 5)	153	170	NNNAATVLQLPQGTTLPK (SEQ ID NO: 10)	94.12	
Sequences were obtained from GenBank (accession number: MN908947 and NC_004718), National Center for Biotechnology Information (NCBI); aa: amino acid, SARS-CoV: Severe Acute Respiratory Syndrome Coronavirus, SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2.								

Table 5: Peptide sequences of hit SARS-CoV-2 peptide pools

Pool	Peptide	Starting aa	Ending aa	Polypeptide sequence	SEQ ID NO:
Spike (S) glycoprotein					
S6	P1	201	218	FKIYSKHTPINLVRDLPPQ	11
	P2	209	226	PINLVRDLPPQGFSALEPL	12
	P3	217	234	PQGFSALEPLVDLPIGIN	13
	P4	225	242	PLVDLPIGINITRFQTLL	14
	P5	233	250	INITRFQTLLALHRSYLT	15
S14	P1	521	538	PATVCGPKKSTNLVKNKC	16
	P2	529	546	KSTNLVKNKCVNFNFNGL	17
	P3	537	554	KCVNFNFNGLTGTGVLTE	18
	P4	545	562	GLTGTGVLTESNKKFLPF	19
	P5	553	570	TESNKKFLPFQQFGRDIA	2
S20	P1	761	778	TQLNRALTGIAVEQDKNT	20
	P2	769	786	GIAVEQDKNTQEVFAQVK	3
	P3	777	794	NTQEVFAQVKQIYKTPPI	21
	P4	785	802	VKQIYKTPPIKDFGGFNF	22

	P5	793	810	PIKDFGGFNFSQILPDPS	23
S21	P1	801	818	NFSQILPDPSKPSKRSFI	24
	P2	809	826	PSKPSKRSFIEDLLFNKV	4
	P3	817	834	FIEDLLFNKVTADAGFI	25
	P4	825	842	KVTADAGFIKQYGDCLG	26
	P5	833	850	FIKQYGDCLGDIAARDLI	27
Nucleocapsid (N) phosphoprotein					
N4	P1	121	138	LPYGANKDGIWVATEGA	28
	P2	129	146	GIIWVATEGALNTPKDH	29
	P3	137	154	GALNTPKDHIGTRNPANN	30
	P4	145	162	HIGTRNPANNAAIVLQLP	31
	P5	153	170	NNAAIVLQLPQGTTLPKG	5
N10	P1	361	378	KTFPPTPEKKDKKKKADE	32
	P2	369	386	KKDKKKKADETQALPQRQ	33
	P3	377	394	DETQALPQRQKKQQTIVL	34
	P4	385	402	RQKKQQTIVLLPAADLDD	35
	P5	393	410	TLLPAADLDDFSKQLQQS	36
	P6	401	418	DDFSKQLQQSMSSADSTQ	37
	P7	402	419	DDFSKQLQQSMSSADSTQA	38
Sequences were obtained from GenBank (accession number: MN908947), National Center for Biotechnology Information (NCBI); aa: amino acid, SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2.					

Putative SARS-CoV-2 epitopes can serve as highly specific and sensitive tools to identify patients exposed to SARS-CoV-2

Anti-SARS-CoV-2 antibody levels differ with infection phases, with IgM being detected as early as 3 days p.i. (23). To investigate whether IgM epitopes could be identified with the same approach, plasma samples collected during the early acute phase of median 5 days p.i. were screened alongside with samples taken at median 14 days p.i. (FIG. 6A). Two positive pools, S20 from S glycoprotein, and N10 from N protein, were detected albeit with a lower recognition level at median 5 days p.i., as compared to samples taken during median 14 days p.i. (FIG. 6A). These peptide pools were further characterised to determine the specific regions. Peptides S20P4 and N10P1 were dominant hits in their respective pools (FIG. 6B and Table 5). Interestingly, COVID-19 patient plasma samples were able to recognise both corresponding hit regions on SARS-CoV S and N proteins, suggesting that these are SARS cross-reactive epitopes for IgM detection (FIG. 6C). Peptide S20P4 is localised in the surface of the S

viral protein (FIG. 6D). However, when validated with a larger set of patient samples, the putative epitopes were only recognised by a minority of the COVID-19 patients, indicating that this approach is inefficient for IgM detection (FIG. 6E and Table 6).

5

Table 6: Receiver operating characteristic (ROC) profiles of IgM-specific SARS-CoV-2 linear B-cell epitopes

Peptide	Threshold	Specificity (%)	Sensitivity (%)	AUC	Wilcoxon P_{adj} value
Median 10 days pio					
S20P4	0.168	30.488	83.051	0.519	0.702
N10P1	0.104	95.122	25.424	0.600	0.0562
Median 23 days pio					
S20P4	0.099	30.488	84.848	0.537	0.4706
N10P1	0.152	73.171	36.363	0.538	0.4706
pio: post-illness onset, AUC: area under curve, SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2.					

To further determine the longitudinal profiling in a larger cohort of COVID-19 patients, individual plasma samples collected at median 10 days pio (n=59) and median 23 days pio (n=66) were assessed and compared. All the four putative epitopes, S14P5, S20P2, S21P2, and N4P5, except S6P2, were distinctly recognised by plasma samples from most COVID-19 patients (FIG. 2A). Strikingly, there was no recognition to all epitopes by samples obtained from patients infected with seasonal hCoV (baseline samples from patients before infection were used as controls), and healthy donors (FIG. 2A). However, cross-reactivity was observed for epitopes S14P5 and S21P2 from one out of 20 recovered SARS individuals (FIG. 2A).

In order to define an optimal cut-off value for each epitope peptide, the Youden J statistical method was used, which determines the sensitivity and specificity profiles of the assay in one single test (17). This will guide the development of diagnostic tests and estimate the probability of an informed decision. Epitopes S20P2 and N4P5 exhibited specificity and sensitivity profiles of about 90% (FIG. 2B and Table 3). Both S14P5 and S21P2 have a moderate specificity and sensitivity level of >80%, while S6P2 has a poor ability to correctly

detect COVID-19 patients with a sensitivity level of <10% (FIG. 2B and Table 3). S14P5 and N4P5 were consistently recognised by COVID-19 patients at a frequency of >97%, whereas the frequency of recognising S20P2 and S21P2 ranged between 81 to 88% at median 10 days pio (FIG. 2C). However, at median 23 days pio, the percentage recognition for all four epitopes is > 95%. Interestingly, when the various peptide combinations of S14P5, S20P2, S21P2 and N4P5 are analysed by predictive bioinformatics tools, specificity and sensitivity of detection could achieve 100% (Table 4).

10 **Table 3: Receiver operating characteristic (ROC) profiles of IgG-specific SARS-CoV-2 linear B-cell epitopes**

Peptide	Threshold	Specificity (%)	Sensitivity (%)	AUC	Wilcoxon P_{adj} value
Median 10 days pio					
S6P2	1.153	100.00	3.390	0.847	<0.0001
S14P5	0.288	87.805	96.610	0.915	<0.0001
S20P2	0.328	95.122	88.136	0.974	<0.0001
S21P2	0.275	76.829	81.356	0.864	<0.0001
N4P5	0.313	100.00	96.610	0.994	<0.0001
Median 23 days pio					
S6P2	0.022	96.341	12.121	0.895	<0.0001
S14P5	0.354	91.463	96.970	0.962	<0.0001
S20P2	0.360	98.780	95.455	0.993	<0.0001
S21P2	0.291	80.488	96.970	0.943	<0.0001
N4P5	0.335	100.00	98.485	0.999	<0.0001
pio: post-illness onset, AUC: area under curve, SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2.					

Table 4: Receiver operating characteristic (ROC) profiles of IgG-specific SARS-CoV-2 linear B-cell epitopes used in combination

Number of peptides in combination	Peptide combination	Threshold	Specificity (%)	Sensitivity (%)	AUC
Median 10 days pio					
4	S14P5 + S20P2 + S21P2 + N4P5	0.500	100.00	100.00	1.000
3	S14P5 + S20P2 + N4P5	0.500	100.00	100.00	1.000

3	S20P2 + S21P2 + N4P5	0.500	100.00	100.00	1.000
3	S14P5 + S21P2 + N4P5	0.804	100.00	98.30	0.997
3	S14P5 + S20P2 + S21P2	0.414	92.70	91.50	0.981
2	S20P2 + N4P5	0.269	98.80	100.00	1.000
2	S14P5 + N4P5	0.803	100.00	98.30	0.997
2	S21P2 + N4P5	0.481	98.80	98.30	0.992
2	S20P2 + S21P2	0.407	92.70	91.50	0.982
2	S14P5 + S20P2	0.503	93.90	88.10	0.975
2	S14P5 + S21P2	0.373	84.10	96.60	0.929
Median 23 days pio					
4	S14P5 + S20P2 + S21P2 + N4P5	0.500	100.00	100.00	1.000
3	S14P5 + S20P2 + N4P5	0.500	100.00	100.00	1.000
3	S20P2 + S21P2 + N4P5	0.500	100.00	100.00	1.000
3	S14P5 + S21P2 + N4P5	0.794	100.00	98.50	0.999
3	S14P5 + S20P2 + S21P2	0.353	98.80	100.00	0.998
2	S20P2 + N4P5	0.500	100.00	100.00	1.000
2	S14P5 + N4P5	0.820	100.00	98.50	0.999
2	S21P2 + N4P5	0.792	100.00	98.50	0.999
2	S20P2 + S21P2	0.307	97.60	97.00	0.998
2	S14P5 + S20P2	0.278	97.60	100.00	0.997
2	S14P5 + S21P2	0.485	92.70	97.00	0.971

pio: post-illness onset, AUC: area under curve, SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2.					

Epitope-specific IgG levels associate with disease severity

SARS-CoV-2 infected patients manifest different disease outcomes that can be categorised into three severity groups: mild (fever and flu-like symptoms only, score 0), moderate (pneumonia with no hypoxia, score 1), or severe (pneumonia with hypoxia, score 2) (24). To elucidate if IgG anti-viral responses at different days pio were associated with disease severity, antibody responses against the four highly recognised epitope peptides were analysed based on disease severity (mild, moderate, and severe) (FIG. 2A and FIG. 3A). Surprisingly, only IgG responses at median 23 days pio against epitopes S14P5, S21P2 and N4P5 were higher in severe COVID-19 patients than in the mild and moderate groups (FIG. 7A and FIG. 3A). It was also observed that IgG responses to epitopes S21P2 and N4P5 were associated with pneumonia at median 23 days pio (FIG. 3B). In more severe cases with patients requiring admission into ICU or oxygen supply, IgG levels to S14P5 at median 23 days pio were also significantly higher (FIG. 3C). Correlation analyses of these epitopes with other clinical parameters of severity showed association with C-reactive protein (CRP), lactate dehydrogenase (LDH) and lymphopenia (FIG. 7B-D).

The development of sensitive and specific immunoassays that can determine SARS-CoV-2 exposure, and serve as sero-epidemiological surveillance tools would greatly help to control virus transmission. In this study, four immunodominant IgG-specific SARS-CoV-2 linear epitopes, S14P5, S20P2, S21P2 and N4P5, located on the S and N viral proteins were identified. These two antigens are commonly used for developing serological assays against coronaviruses (5, 8, 20). Sequences on these epitopes are well conserved across multiple strains of the SARS-CoV-2 (18) (Table 7). The best performing epitopes, S20P2 and N4P5 could correctly discriminate SARS-CoV-2 from seasonal CoV infection in patients, with specificity and sensitivity levels of >90% for single peptide detection.

Table 7: List of mutations from 10956 SARS-CoV-2 sequences curated by China National Center for Bioinformation (CNCB)

Peptide	Genome Position	Virus Number (with variation)	Annotation Type	Protein Position, Amino Acids Change	Gene Position, Codons	Impact Ensembl Variation - Calculated Variant Consequences	Last Updated
S20P2	28368	1	missense_variant	QHD43423.2p.169C>V	gene-S.c.2306Gca>gTa	MODERATE	2020-04-20, 01:13:15
	28376	2	missense_variant	QHD43423.2p.772V>I	gene-S.c.2314Gac>Att	MODERATE	2020-04-20, 01:13:15
	28893	1	synonymous_variant	QHD43423.2p.777N	gene-S.c.2331aat>aat	LOW	2020-04-20, 01:13:15
	23900	1	missense_variant	QHD43423.2p.788E>Q	gene-S.c.2338Gac>Caa	MODERATE	2020-04-20, 01:13:16
	28732	1	synonymous_variant	QHD43423.2p.153N	gene-N.c.439aaC>aat	LOW	2020-04-20, 01:13:33
	28737	4	missense_variant	QHD43423.2p.155A>V	gene-N.c.464gC>gTt	MODERATE	2020-04-20, 01:13:33
N4PS	28739	6	missense_variant	QHD43423.2p.156A>S	gene-N.c.466Gca>Tca	MODERATE	2020-04-20, 01:13:33
	28743	1	missense_variant	QHD43423.2p.157I>T	gene-N.c.470Tca>aCc	MODERATE	2020-04-20, 01:13:33
	28744	2	synonymous_variant, coding sequence variant	QHD43423.2p.157I; QHD43423.2p.157-	gene-N.c.471atC>aTt; gene-N.c.471atC>aY	LOW, MODIFIER	2020-04-20, 01:13:33
	28747	1	synonymous_variant	QHD43423.2p.138V	gene-N.c.474gG>gTt	LOW	2020-04-20, 01:13:33
	28754	1	missense_variant	QHD43423.2p.161L>F	gene-N.c.481Ctt>Ttt	MODERATE	2020-04-20, 01:13:33
	28760	1	missense_variant	QHD43423.2p.163Q>X	gene-N.c.487Caa>Aaa	MODERATE	2020-04-20, 01:13:33
	28762	1	synonymous_variant	QHD43423.2p.163Q	gene-N.c.489caaA>caG	LOW	2020-04-20, 01:13:33
	28765	1	synonymous_variant	QHD43423.2p.164G	gene-N.c.492gga>ggG	LOW	2020-04-20, 01:13:33
	28775	1	missense_variant	QHD43423.2p.168P>S	gene-N.c.502Cca>Tca	MODERATE	2020-04-20, 01:13:33

Interestingly, immunoassays using a combination of two or more epitopes can achieve 100% sensitivity and specificity for the detection of COVID-19 patients. This is an important aspect for the design of point-of-care tests (POCTs) where the combination of two peptides (i.e. S20P2 plus N4P5) could allow for rapid IgG detection with a high level of specificity and sensitivity. This remains as one of the current limitations in the development of POCTs based on single SARS-CoV-2 proteins. Furthermore, the use of recombinant proteins may lead to false positive results and hamper specificity of the immune-based assays due to a high level of cross-reactivities between SARS-CoV-2 and other human coronaviruses such as the closely-related SARS-CoV (11, 22, 25). The prevalence of human coronaviruses differ geographically and in different populations (9, 26). As a result, this will influence the sensitivity of the immune-based assays due to the different cut-off values taken from various healthy controls around the world. However, the use of specific peptide epitopes reported in this study may overcome the various specificity and sensitivity issues.

The wide spectrum of COVID-19 clinical severity ranges from asymptomatic to febrile with mild respiratory tract distress, to viral pneumonia, and to death from respiratory failure or associated complications (8, 23, 27, 28). Our study revealed that IgG levels against epitopes S14P5, S21P2 and N4P5 in COVID-19 patients were associated with disease severity, confirming recent studies that reported an association of antibody levels against the S proteins with disease severity (3, 5, 7, 24). Moreover, within the severe group, high antibody levels against these linear B-cell epitopes were also detected by patients who either developed pneumonia or were admitted into the ICU. Antibody levels against these epitopes also correlated with clinical laboratory parameters such as CRP and LDH levels, which are elevated in severe COVID-19 patients (29, 30). High IgG levels against epitope N4P5 correlated with lymphopenia, a phenomenon found in COVID-19 patients (31). One plausible hypothesis for these associations is that severe patients had a higher viral load during the early infection phase that resulted in higher antibody response.

Altogether, in view of the recent studies demonstrating the presence of mutation and deletions in SARS-CoV-2, the identification of multiple putative epitopes for the detection of SARS-CoV-2 infection allows for constant refinement of the peptide sequences to be used in serological assays (32, 33). An immediate application of these peptide epitopes could be developed as paper-based lateral flow assays (23). Ideally, the implementation of peptides on POCTs would be easier than protein-based systems. Furthermore, the identification of IgG-specific epitopes on the S and N viral proteins expands the toolbox needed for sero-epidemiological studies and vaccine assessment.

Identification of new potential peptides recognised by COVID-19 patients

SARS-CoV-2 linear peptides of the spike protein were synthesised based on GenBank accession number MN908947.3 (Mimotopes). Peptides were biotinylated and were 18-mer in length, with 10 overlapping residues. Peptides were used individually or as 31 pooled sets of five peptides.

Briefly, the peptide library was screened in a peptide-based ELISA assay using pooled plasma samples from COVID-19 patients at 1, 3 and 6 months post-illness onset (pio). COVID-19 patients were stratified into low and high binders based on the anti-SARS-CoV-2 spike IgG levels (<50% and >50% IgG recognition, respectively) (34). Recovered SARS patients and two groups of healthy donors (old, median age 63 years; young, median age 33 years) were also screened in parallel as negative controls.

The preliminary screening revealed several regions of interest (FIG. 8). Firstly, pooled peptides 17, 21 and 29 were recognised by both low and high binders groups of COVID-19 patients, although the latter showed a stronger recognition (FIG. 8A and B). Previously, the present inventors had identified three immunodominant regions in the spike protein, pooled peptides 14, 20 and 21, which were strongly detected during the early SARS-CoV-2 infection phase (35). Interestingly, pooled peptides 14 and 20 were weakly recognised by COVID-19 patients at 3 and 6 months pio (FIG. 8A and B). Taken together, these observations pushed for an investigation into pooled peptides 17 and 29 to determine the specific SARS-CoV-2 immunodominant epitope within these pools.

Identification of specific SARS-CoV-2 immunodominant epitopes on spike pooled peptides 17 and 29

In order to determine the specific region within the pooled peptide, a peptide-based ELISA of the individual peptides of pools 17 (FIG. 9) and 29 (FIG. 10) were carried out, with both pooled and individual patient plasma samples.

Peptides 2, 4 and 5 of pool 17 (herein known as S17P2, S17P4 and S17P5) were highly recognised by the COVID-19 high binders group across the timepoints (FIG. 9A). Although COVID-19 patients in the low binders group showed an IgG response against S17P2, there were also detectable responses from the recovered SARS and healthy controls (FIG. 9A). Nonetheless, S17P2, together with S17P4 and S17P5, were further validated by testing against individual plasma samples (FIG. 9B). However, all or majority of COVID-19 patients showed IgG levels below the detection cut-off (mean + 3SD of healthy controls) against the three peptides (FIG. 9B). The strong response observed from COVID-19 high binders groups against S17P4 and S17P5 in Figure 2A was likely the effect from a single patient (FIG. 9B). Therefore, S17 was not pursued further.

In contrast, peptide 4 of pool 29 (herein known as S29P4) on SARS-CoV-2 was highly recognised by IgG antibodies in the pooled plasma samples of COVID-19 patients, particularly from the high binders group (FIG. 10A). There were no recognition of peptides S29P1 to S29P3 and S29P5 among the COVID-19 patients (FIG. 10A). When S29P4 was further validated with individual plasma samples, IgG responses from several COVID-19 patients across the different timepoints were observed above the detection cut-off (mean + 3SD of healthy controls). Notably, none of the recovered SARS subjects had positive IgG recognition against this peptide (FIG. 10B), despite the sequence of S29P4 showing 100% amino acid identity with that of SARS-CoV (GenBank accession number NC_004718.3).

Interestingly, using a prefusion structure of the SARS-CoV-2 spike protein, S29P4 is localised away from the receptor binding domain (RBD) (FIG. 10C).

Additionally, it is also hidden in the closed conformation, and only appears when the spike protein is primed to bind to the ACE2 receptor (FIG. 10D). The mechanisms behind these conformational changes still remain unknown. Nevertheless, S29P4 could be used to detect for SARS-CoV-2 exposure in COVID-19 patients.

The sequences of peptides within pools 17 and 29 are listed in Table 1.

Table 8: Sequences of SARS-CoV-2 peptides in pool 17 and 29

Pooled Peptide	Peptide	Antigen	Virus	Amino acid residues	Peptide sequence (linker)	SEQ ID NO:
17	S17P1	Spike	SARS-CoV-2	641-658	SGSGNVFQTRA GCLIGAEHVNN	39
	S17P2	Spike	SARS-CoV-2	649-666	SGSGCLIGAEHV NNSYECDIPI	40
	S17P3	Spike	SARS-CoV-2	657-674	SGSGNNSYECD IPIGAGICASY	41
	S17P4	Spike	SARS-CoV-2	665-682	SGSGPIGAGICA SYQTQTNSPR	42
	S17P5	Spike	SARS-CoV-2	673-690	SGSGSYQTQTN SPRRARSVASQ	43
29	S29P1	Spike	SARS-CoV-2	1121-1138	SGSGFVSGNCD VVIGIVNNTVY	44
	S29P2	Spike	SARS-CoV-2	1129-1146	SGSGVIGIVNNT VYDPLQPELD	45
	S29P3	Spike	SARS-CoV-2	1137-1154	SGSGVYDPLQP ELDSFKEELDK	46
	S29P4	Spike	SARS-CoV-2	1145-1162	SGSGLDSFKEEL DKYFKNHTSP	47
	S29P5	Spike	SARS-CoV-2	1153-1170	SGSGDKYFKNH TSPDVLGDIS	48

Validation of S29P4 as an immunodominant SARS-CoV-2 epitope

To further validate S29P4 as a long-lived antibody-binding epitope, the IgG levels of >100 COVID-19 patients at 1, 3 and 6 months p.o. were determined. The IgG response from COVID-19 patients against S29P4 was the highest at 1 month, and declined at 6 months p.o. (FIG. 11A). To evaluate the specificity and sensitivity of this epitope, a receiver operating characteristic (ROC) profile was assessed (FIG. 11B). At 1 month p.o., S29P4 was shown to have an area under

curve (AUC) of 0.925, and specificity and sensitivity of 97.6% and 77.8%, respectively (FIG. 11B).

Comparison of S29P4 with previously characterised immunodominant epitopes

In order to assess the longitudinal profiles of previously characterised epitopes S14P5, S20P2, S21P2 and N4P5, COVID-19 plasma samples at 1, 3 and 6 months p.i. were tested similarly as above. The performance of these four epitopes and S29P4 was then compared.

The sequences of S14P5, S20P2, S21P2 and N4P5 are shown in Table 9.

Table 9: Sequences of previously characterised SARS-CoV-2 epitopes S14P5, S20P2, S21P2 and N4P5

Peptide	Antigen	Virus	Amino acid residues	Peptide sequence	SEQ ID NO:
S14P5	Spike	SARS-CoV-2	553-570	TESNKKFLPFQQFGRDIA	2
S20P2	Spike	SARS-CoV-2	769-792	GIAVEQDKNTQEVFAQVKQIYKTP	49
S21P2	Spike	SARS-CoV-2	809-826	PSKPSKRSFIEDL LFNKV	4
N4P5	Nucleo-capsid	SARS-CoV-2	153-174	NNAAIVLQLPQGT TLPKGFYAE	50

As observed in S29P4 (FIG. 11A), the IgG levels from COVID-19 patients against S14P5, S20P2, S21P2 and N4P5 declined from 1 to 6 months p.i. (FIG. 12A). At 1 month and 3 months p.i., S29P4 showed the highest positive recognition at 73.6% and 59.4%, respectively (FIG. 12B). At 6 months p.i., S21P2 and S29P4 had similar levels of recognition, at 35.9% and 34.0%, respectively (FIG. 12B).

In conclusion, the present disclosure identifies another immunodominant linear B cell epitope on the SARS-CoV-2 spike glycoprotein, S29P4 (amino acid residues 1145-1162). The inclusion of S29P4 to the four previously reported

immunodominant epitopes may provide greater coverage to detect for COVID-19 exposure in the population, and may also serve as a potential vaccine candidate.

Identification of potential peptides recognized by COVID-19 patients

5 SARS-CoV-2 and SARS-CoV linear peptides of the spike protein were synthesized based on GenBank accession numbers MN908947.3 and NC_004718.3, respectively (Mimotopes). Peptides were biotinylated and were 18-mer in length, with 10 overlapping residues. Peptides were used either individually or as pooled sets of five peptides.

10

The peptide libraries were screened in a peptide-based ELISA assay using serum samples from recovered COVID-19 patients, and healthy recalled subjects that had a previous SARS infection during the Singapore outbreak 17 years ago. Serum samples were first screened with pooled peptides of spike SARS-CoV-2 and SARS-CoV to identify potential areas of interest (FIG. 13).

15 The preliminary screening revealed two distinct regions of interest (FIG. 13). Pooled peptide 14 of SARS-CoV-2 was recognized strongly by COVID-19 patients (FIG. 13A). However, for the same patients, there were no recognition of pools 44 and 45, the corresponding pooled peptides of this region on SARS-CoV (FIG. 13B). Interestingly, the recovered SARS subjects had low peptide recognition on these pooled peptides, suggesting that pooled peptide 14 (amino acid residues 530-579 on SARS-CoV-2) could potentially be a specific SARS-CoV-2 antigenic target.

20 Furthermore, there was a strong recognition on pooled peptide 21 on SARS-CoV-2 (FIG. 13A) and the corresponding pool 51 (amino acid residues 761-810) on SARS-CoV (FIG. 13B) by the COVID-19 patients. Recovered SARS subjects however, did not show a positive recognition to these pooled peptides (FIG. 13). Taken together, these observations suggest that this region encompassing amino acid residues 801-850 on SARS-CoV-2 could be a pan-coronavirus target.

25

30

Identification of a specific SARS-CoV-2 immunodominant epitope on spike protein

In order to determine the specific region within the pooled peptide, a peptide-based ELISA of the individual peptides of pool 14 was carried out.

Peptide 14-5 of pool 14 on SARS-CoV-2 was highly recognized by IgG antibodies in the serum samples of recovered COVID-19 patients tested (FIG. 14A). The remaining peptides (14-1 to 14-4) within this pool however, were not recognized among the patients (FIG. 14A). To determine the specificity of the region, the sera of COVID-19 patients were also tested on peptide 45-3, the corresponding region on SARS-CoV. There was a significant reduction in binding on the SARS-CoV peptide 45-3 (FIG. 14B). Using a prefusion structure of the SARS-CoV-2 spike protein, peptide 14-5 is localized in proximity of the receptor binding domain (RBD) (FIG. 14C). Thus, it is plausible that antibodies binding to this region may sterically hinder binding to the host angiotensin-converting enzyme 2 (ACE2).

Identification of a pan-coronavirus immunodominant epitope on spike protein

Peptide-based ELISA for the individual peptides of spike SARS-CoV-2 pool 21 was similarly carried out.

IgG antibodies in the convalescent sera of recovered COVID-19 patients was able to strongly recognize peptide 21-2 within the pool 21 of SARS-CoV-2 (FIG. 15A). Interestingly, this peptide contains a part of the fusion peptide sequence, which may potentially affect virus fusion (FIG. 15C).

The sequences of relevant peptides are detailed in Table 10.

Table 10. Sequences of SARS-CoV and SARS-CoV-2 peptides

Pooled Peptide	Peptide	Antigen	Virus	Amino acid residues	Peptide sequence (linker)	SEQ ID NO:
	14-1	Spike	SARS-CoV-2	521-538	SGSGPATVCGPKKSTNLVKNKC	51
	14-2	Spike	SARS-CoV-2	529-546	SGSGKSTNLVKNKCVNFNFNGL	52

14	14-3	Spike	SARS-CoV-2	537-554	SGSGKCVNFNFNGLT GTGVLTE	53
	14-4	Spike	SARS-CoV-2	545-562	SGSGGLTGTGVLTES NKKFLPF	54
	14-5	Spike	SARS-CoV-2	553-570	SGSGTESNKKFLPFQ QFGRDIA	55
21	21-1	Spike	SARS-CoV-2	801-818	SGSGNFSQILPDPSK PSKRSEI	56
	21-2	Spike	SARS-CoV-2	809-826	SGSGPSKPSKRSEI DLLFNKV	57
	21-3	Spike	SARS-CoV-2	817-834	SGSGFIEDLLFNKVTL ADAGFI	58
	21-4	Spike	SARS-CoV-2	825-842	SGSGKVTADAGFIK QYGDCLG	59
	21-5	Spike	SARS-CoV-2	833-850	SGSGFIKQYGDCLGDI AARDLI	60
51	51-1	Spike	SARS-CoV	761-778	SGSGREVFAQVKQM YKTPTLKY	61
	51-2	Spike	SARS-CoV	769-786	SGSGQMYKTPTLKYF GGFNFSQ	62
	51-3	Spike	SARS-CoV	777-794	SGSGKYFGGFNFSQI LPDPLKP	63
	51-4	Spike	SARS-CoV	785-802	SGSGSQILPDPLKPTK RSFIED	64
	51-5	Spike	SARS-CoV	793-810	SGSGKPTKRSFIEDLL FNKVTL	65
45	45-3	Spike	SARS-CoV	537-554	SGSGVLTSSKRFQP FQQFGRD	66

Epitope derivatives and modifications to sequences

In order to improve the detection of SARS-CoV-2 antibodies, different lengths of the immunodominant epitopes of S14P5, S20P2, S21P2 and N4P5 were designed (FIG. 16B). Pooled plasma samples of mild, moderate and severe COVID-19 patients at convalescent and recovered phases, as well as recovered SARS patients, and old and young healthy donors were tested in a peptide-based ELISA against the different peptide derivatives (FIG. 16B). The original versions of S14P5 and S21P2 demonstrated the best detection profiles, whereas higher signals were observed in the different COVID-19 groups against the extended sequences of S20P2 and N4P5 (FIG. 16A).

References

1. Cohen J, Normile D. New SARS-like virus in China triggers alarm. *Science*. 2020;367(6475):234.
2. WHO. Coronavirus disease (COVID-2019) situation reports [Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>.]
3. Long Q-x, Deng H-j, Chen J, Hu J, Liu B-z, Liao P, et al. Antibody responses to SARS-CoV-2 in COVID-19 patients: the perspective application of serological tests in clinical practice. *medRxiv* [Preprint]. 2020:2020.03.18.20038018.
4. Amanat F, Nguyen T, Chromikova V, Strohmeier S, Stadlbauer D, Javier A, et al. A serological assay to detect SARS-CoV-2 seroconversion in humans *medRxiv* [Preprint]. 2020:2020.03.17.20037713.
5. Okba NMA, Muller MA, Li W, Wang C, GeurtsvanKessel CH, Corman VM, et al. SARS-CoV-2 specific antibody responses in COVID-19 patients. *Emerg Infect Dis*. 2020;26(7).
6. Xiao DAT, Gao DC, Zhang DS. Profile of Specific Antibodies to SARS-CoV-2: The First Report. *J Infect*. 2020:S0163-4453.
7. Zhang B, Zhou X, Zhu C, Feng F, Qiu Y, Feng J, et al. Immune phenotyping based on neutrophil-to-lymphocyte ratio and IgG predicts disease severity and outcome for patients with COVID-19. *medRxiv* [Preprint]. 2020:2020.03.12.20035048.
8. Ahmed SF, Quadeer AA, McKay MR. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses*. 2020;12(3).
9. Gaunt ER, Hardie A, Claas ECJ, Simmonds P, Templeton KE. Epidemiology and Clinical Presentations of the Four Human Coronaviruses 229E, HKU1, NL63, and OC43 Detected over 3 Years Using a Novel Multiplex Real-Time PCR Method. *Journal of Clinical Microbiology*. 2010;48(8):2940.
10. Parashar UD, Anderson LJ. Severe acute respiratory syndrome: review and lessons of the 2003 outbreak. *International Journal of Epidemiology*. 2004;33(4):628-34.

11. Che XY, Qiu LW, Liao ZY, Wang YD, Wen K, Pan YX, et al. Antigenic cross-reactivity between severe acute respiratory syndrome-associated coronavirus and human coronaviruses 229E and OC43. *J Infect Dis.* 2005;191(12):2033-7.
- 5 12. Pung R, Chiew CJ, Young BE, Chin S, Chen MI, Clapham HE, et al. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. *Lancet.* 2020;395(10229):1039-46.
13. Ang LW, Kam YW, Lin C, Krishnan PU, Tay J, Ng LC, et al. Seroprevalence of antibodies against chikungunya virus in Singapore resident
10 adult population. *PLoS neglected tropical diseases.* 2017;11(12):e0006163-e.
14. Darnell MER, Taylor DR. Evaluation of inactivation methods for severe acute respiratory syndrome coronavirus in noncellular blood products. *Transfusion.* 2006;46(10):1770-7.
15. Kam Y-W, Lum F-M, Teo T-H, Lee WWL, Simarmata D, Harjanto S, et al. Early neutralizing IgG response to Chikungunya virus in infected patients targets
15 a dominant linear epitope on the E2 glycoprotein. *EMBO molecular medicine.* 2012;4(4):330-43.
16. Kam YW, Lee CY, Teo TH, Howland SW, Amrun SN, Lum FM, et al. Cross-reactive dengue human monoclonal antibody prevents severe pathologies and
20 death from Zika virus infections. *JCI Insight.* 2017;2(8).
17. Kam YW, Leite JA, Amrun SN, Lum FM, Yee WX, Bakar FA, et al. ZIKV-Specific NS1 Epitopes as Serological Markers of Acute Zika Virus Infection. *J Infect Dis.* 2019;220(2):203-12.
18. Poh CM, Carissimo G, Wang B, Amrun SN, Lee CY-P, Chee RS-L, et al.
25 Potent neutralizing antibodies in the sera of convalescent COVID-19 patients are directed against conserved linear epitopes on the SARS-CoV-2 spike protein. *bioRxiv.* 2020:2020.03.30.015461.
19. Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950;3(1):32-5.
20. Lin Y, Shen X, Yang RF, Li YX, Ji YY, He YY, et al. Identification of an
30 epitope of SARS-coronavirus nucleocapsid protein. *Cell Res.* 2003;13(3):141-5.
21. He Y, Li J, Heck S, Lustigman S, Jiang S. Antigenic and Immunogenic Characterization of Recombinant Baculovirus-Expressed Severe Acute

- Respiratory Syndrome Coronavirus Spike Protein: Implication for Vaccine Design. *Journal of Virology*. 2006;80(12):5757.
22. Lv H, Wu NC, Tsang OT-Y, Yuan M, Perera RAPM, Leung WS, et al. Cross-reactive antibody response between SARS-CoV-2 and SARS-CoV infections. *bioRxiv [Preprint]*. 2020:2020.03.15.993097.
23. Lee CY-P, Lin RTP, Renia L, Ng LFP. Serological Approaches for COVID-19: Epidemiologic Perspective on Surveillance and Control. *Frontiers in Immunology*. 2020;11(879).
24. Wong JEL, Leo YS, Tan CC. COVID-19 in Singapore—Current Experience: Critical Global Issues That Require Attention and Action. *JAMA*. 2020;323(13):1243-4.
25. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature Communications*. 2020;11(1):1620.
26. van der Hoek L. Human coronaviruses: what do they cause? *Antivir Ther*. 2007;12(4 Pt B):651-8.
27. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395(10223):507-13.
28. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *Jama*. 2020.
29. Gao Y, Li T, Han M, Li X, Wu D, Xu Y, et al. Diagnostic utility of clinical laboratory data determinations for patients with the severe COVID-19. *J Med Virol*. 2020.
30. Mardani R, Ahmadi Vasmehjani A, Zali F, Gholami A, Mousavi Nasab SD, Kaghazian H, et al. Laboratory Parameters in Detection of COVID-19 Patients with Positive RT-PCR; a Diagnostic Accuracy Study. *Archives of academic emergency medicine*. 2020;8(1):e43-e.
31. Fan BE, Chong VCL, Chan SSW, Lim GH, Lim KGE, Tan GB, et al. Hematologic parameters in patients with COVID-19 infection. *Am J Hematol*. 2020.

32. Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, et al. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J Med Virol.* 2020.
33. Phan T. Genetic diversity and evolution of SARS-CoV-2. *Infect Genet Evol.* 2020;81:104260.
- 5 34. Goh YS, Chavatte J, Lim AJL, Lee B, Amrun SN, Lee CYP, et al. Sensitive Detection of Total Anti-Spike Antibodies and Isotype Switching in Asymptomatic and Symptomatic COVID-19 Patients. Manuscript accepted for publication at *Cell Reports Medicine*.
35. Amrun SN, Lee CYP, Lee B, Fong SW, Young BE, Chee RSL, et al. Linear
10 B-cell epitopes in the spike and nucleocapsid proteins as markers of SARS-CoV-2 exposure and disease severity. *EBiomedicine* 2020; 58:102911.

APPLICATIONS

Embodiments of the present disclosure provides epitopes that can
15 generate antigen binding molecules (such as antibodies) that serve as useful indicators for recent or past SARS-CoV-2 infections. Also disclosed are antibodies that can function as sero-immunosurveillance tools with high specificity and sensitivity.

The epitopes of the present disclosure are advantageously flexible and
20 can be used alone or in combination to allow for the development of improved point-of-care-tests (POCTs).

It will be appreciated by a person skilled in the art that other variations and/or modifications may be made to the embodiments disclosed herein without departing from the spirit or scope of the disclosure as broadly
25 described. For example, in the description herein, features of different exemplary embodiments may be mixed, combined, interchanged, incorporated, adopted, modified, included etc. or the like across different exemplary embodiments. The present embodiments are, therefore, to be considered in all respects to be illustrative and not restrictive.