

USE OF HLA-A\*11:01-RESTRICTED HEPATITIS B VIRUS (HBV) PEPTIDES  
FOR IDENTIFYING HBV-SPECIFIC CD8+ T CELLS

The present invention relates to peptides and their ability to identify and bind to T cells specific for HBV-infected hepatocytes.

5

The listing or discussion of an apparently prior-published document in this specification should not necessarily be taken as an acknowledgement that the document is part of the state of the art or is common general knowledge.

10 Any document referred to herein is hereby incorporated by reference in its entirety.

Chronic HBV infection (CHB) remains a major health issue and is the leading causative agent of hepatocellular carcinoma (HCC) worldwide. Despite an effective vaccine, CHB has no cure and many patients are only diagnosed during the later stages of the disease where treatment efficacy is limited. The estimated mortality due to viral hepatitis has escalated by more than 50% in last decade. CHB develops into a series of stages that are defined by a few clinical parameters with limited associated immunological evidence. How adaptive immunity changes as young patients progress from the Immune Tolerant stage (IT, or HBeAg+ chronic infection, high viremia but limited liver inflammation), to progressive Immune Active stage (IA, or HBeAg+ chronic hepatitis, high viremia and high liver inflammation), and for some, spontaneously become HBeAg- Inactive Carriers (InA, or HBeAg- chronic infection, low-to-undetectable viral load and limited liver inflammation), is not thoroughly understood. Some argue that these definitions could affect early treatment opportunity and should be revisited.

25

Despite their low frequency in most CHB patients, the virus-specific T cell response has been of much interest for HBV immunologists. Seminal experiments on CHB patients and animal models such as HBV challenged chimpanzee has shown the indispensable function of virus-specific CD8+ T cells in viral control. Historically, mapping for potential virus-specific CD8+ T cells against HBV has mainly focused on HLA-A\*02:01-restricted epitopes. For instance, numerous studies on a single epitope targeted by A\*02:01-restricted HBVcore18-27-specific CD8+ T cells has provided many implications for immunotherapy. However, in Asia, which

30

has high prevalence of HBV infection, the predominant allele in common East-Asian ethnicities is A\*11:01, whose immunogenicity against chronic HBV is poorly defined. Hence, there is an unmet need to investigate A\*11:01-restricted HBV-specific CD8<sup>+</sup> T cells in CHB. Furthermore, regardless of MHC-restriction, due to the very low frequencies of HBV-specific T cells, information about their un-manipulated phenotypes are lacking.

In murine chronic lymphocytic choriomeningitis virus (LCMV) infection, prolonged and elevated viral antigenic exposure coincides with the upregulation of multiple inhibitory receptors on virus-specific T cells has led to the definition of the state of T cell exhaustion. The resemblance of exhausted T cells (TEX) observed in human chronic viral infection provides an explanation for the functional failure of immune response, but also pinpoints a valuable target to boost host immunity. Evidences suggest that such TEX cells arise from an altered path of memory T cell development. Besides several well-described defects, a hallmark of T cell exhaustion is a progressive loss of functional capacity, which correlates with the cumulative expression of inhibitory receptors over the course of persistent antigen stimulation. This functionally impaired T cell subset has been described in HIV and HCV infection, however, results from other studies do not necessary fit this model. Therefore, by simultaneously measuring a wide range of inhibitory receptors and memory-associated markers, we aimed to evaluate the extent to which the profiles of HBV-specific T cells fit with such model of “Hierarchical T cell exhaustion” in human CHB.

Here, to overcome challenges associated with identifying and deep-profiling un-manipulated HBV-specific T cells, mass cytometry together with a highly multiplexed combinatorial peptide-MHC (pMHC) tetramer strategy was used to simultaneously screen for and interrogate 562 A\*11:01-restricted T cell candidate epitopes. Using a self-validated automatic tetramer deconvolution and unsupervised high-dimensional analyses, it is found that virus-specific CD8<sup>+</sup> T cells targeting HBVpol387 and HBVcore169 displayed complex phenotypic profiles and T cell receptor sequence usages that co-varied with the HBV infection status. Based on a high-dimensional trajectory analysis, it was also found that the profiles of HBV-specific T cells from blood are indicative of the degree of viral control in patients from two separately analyzed cohorts.

The present inventors have now identified amino acid sequences and compositions of peptides and its variants derived from HBVcore169 (also known as HBVcore141) that are restricted by HLA-A1101. Such epitopes can be used to induce cellular response and elicit multifunctional anti-viral T cell activity against HBV. In particular, the present invention  
5 discloses isolated peptides comprising seven oligopeptides, including STLPETAVVRR, STLPETAVVR, STLPETTVVRR, STLPETTVIRR, STPPETTVVRR, STLPETTVVGR, STIPETTVVRR that can be used as antigens in epitope-based therapeutics/vaccines or immunotherapy to prevent and/or treat a Hepatitis B virus (HBV) infection in a patient.

10 In addition, the present inventors have also identified the amino acid and nucleotide sequences comprising four HLA-A\*11:01-restricted epitope-reactive T cell receptors (TCRs) that are specific for HBVcore169 - CASGDSNSPLHF, CASSGGQIVYEQYF, CSARGGRGGDYTF and CASSQDWTEAFF. Such TCR sequences can be constructed and encoded onto T cells to target HBV-infected hepatocytes using adoptive transfer immunotherapy for HLA-A11-1  
15 positive patients.

The general structure of T cell receptors (TCRs), their domain structure and the organisation of genes that encode them is well known, for example see Chapter 11 in Immunology, second edition (1994), by Janis Kuby, W H Freeman & Co, New York, USA, and Garcia et al  
20 (1999) Ann. Rev. Immunol. 17, 369-397. One common class of natural TCRs is the  $\alpha\beta$  class in which the TCRs are made up of a separate alpha chain and a separate beta chain which form a heterodimer which is T cell membrane associated. Each alpha and beta chain is made up of regions which, in order from the N terminus to the C terminus are a leader sequence, a variable region, a constant region, a connecting sequence, a transmembrane region and a  
25 cytoplasmic tail region. The variable region of the alpha chain is called the V $\alpha$  region and the variable region of the beta chain is called the V $\beta$  region. Similarly, the constant region of the alpha chain is called the C $\alpha$  region and the constant region of the beta chain is called the C $\beta$  region. The job of the  $\alpha\beta$  TCR is to recognise and bind to a peptide presented in a HLA molecule of a cell in the body. Generally speaking, the TCR cannot recognise and bind the  
30 peptide unless it is presented by a particular HLA molecule, and the TCR cannot recognise a HLA molecule unless it is presenting the specific peptide. T cells harboring a specific TCR will

target cells which are presenting a specific peptide in a particular HLA molecule on a cell (ie a peptide-HLA complex), and this is the main principle of T cell-based immunity.

The peptide-HLA complex is recognised by the combined V regions of the alpha and beta chains of the TCR. In particular, it is the complementarity determining regions (CDRs) of the V regions which mediate recognition of the peptide-HLA complex. The V region of the alpha and beta chains of the natural TCR are made up of, in order in an N-terminal to C-terminal direction, FR1, CDR1, FR2, CDR2, FR3 and CDR3, where FR stands for "framework region" and CDR stands for "complementarity determining region". The FRs and CDRs of the alpha and beta chains are different. Of note, the CDR3 of beta chain was encoded by V(D)J segment, giving it the higher sequence diversity than alpha chain. It is also well-known that the CDR3 of beta chain is the key determinant for TCR to recognize peptide-MHC complex.

In a first aspect of the invention, there is provided a peptide comprising an amino sequence selected from the group consisting of STLPETAVVRR, STLPETAVVR, STLPETTVVRR, STLPETTVIRR, STPPETTVVRR, STLPETTVVGR and STIPETTVVRR, wherein the peptide is derived from Hepatitis B virus core169 and is capable of binding HLA-A\*1101 and when bound to HLA-A\*1101 is capable of identifying T cells specific for Hepatitis B virus.

Whether or not a peptide binds to HLA-DRB1\*04 may be determined using any method known in the art.

The recognition of harmful pathogens or disease causing mutations within self tissue occurs through two mechanisms within the human immune system. Antibody molecules expressed by B cells bind with biological molecules, typically expressed on the surface of invading microorganisms or deviant self cells, in highly specific manner and will label these molecules in a manner which will trigger an appropriate immune response. In addition to the antibody response, pathogens and disease causing mutations can be detected due to unique proteins expressed by the pathogens or mutations. These proteins are broken down into small peptide fragments by natural and continuous protein degradation systems in the human cell. The peptide fragments will bind with special molecules, referred to as MHC Class I and MHC Class II molecules, expressed on the surface of all cells. In the human, these molecules

are referred to as HLA molecules and are numbered according the large number of alleles which exist across the human population.

In humans, the peptide fragments derived from pathogens bind with specific HLA molecules and are transported to the surface of the cell. The peptides are captured in a particular configuration which allows the peptides to be detected by T cell receptors ("TCRs") expressed on the surface of T cells. Through natural selection and development processes which are linked to the immune system's ability to detect danger signals in connection with the presence of a foreign organism, the human body produces T cells with TCRs which can recognize and distinguish between peptide fragments derived from a harmful pathogen and peptides which are derived from harmless microorganisms or healthy self tissue.

MHC Class I molecules present peptides derived mainly from proteins found within the cell to CD8+ T cells, also referred to as cytotoxic T cells or CTL. The peptides which bind with MHC Class I molecules are usually 8-10 amino acids in length. MHC Class II molecules present peptides derived from proteins or organisms which have been endocytosed from the extracellular milieu. MHC Class II molecules present peptides to CD4+ T cells, also referred to as T helper cells although CD4+ T cells may also have direct cytotoxic functions. The peptides which bind to MHC Class II molecules are relatively unconstrained in terms of length, although Class II peptides generally fall within a range of 13 -17 amino acids, for example 14 or 15 or 16 amino acids.

A particular advantage of a peptide of the invention, is that it binds to the HLA-A\*1101 molecule which is presented on a HLA allele that appears to be prevalent across the Asian patient population, a demographic that is not researched enough.

By "peptide" we include not only molecules in which amino acid residues are joined by peptide (-CO-NH-) linkages but also molecules in which the peptide bond is reversed. Such retro-inverso peptidomimetics may be made using methods known in the art, for example such as those described in Mézière et al (1997) J. Immunol. 159, 3230-3237. This approach involves making pseudopeptides containing changes involving the backbone, and not the orientation of side chains. Mézière et al (1997) show that, at least for MHC class II and T

helper cell responses, these pseudopeptides are useful. Retro-inverse peptides, which contain NH-CO bonds instead of CO-NH peptide bonds, are much more resistant to proteolysis. We also include any protein or polypeptide that by virtue of its origin or source of derivation is not associated with naturally- associated components that accompany it in its native state; is substantially free of other proteins from the same source. A protein may be rendered substantially free of naturally associated components or substantially purified by isolation, using protein purification techniques known in the art. The peptides / proteins obtained in the invention may be "substantially purified" which is meant the protein is substantially free of contaminating agents, e.g., at least about 70% or 75% or 80% or 85% or 90% or 95% or 96% or 97% or 98% or 99% free of contaminating agents.

Similarly, the peptide bond may be dispensed with altogether provided that an appropriate linker moiety which retains the spacing between the C $\alpha$  atoms of the amino acid residues is used; it is particularly preferred if the linker moiety has substantially the same charge distribution and substantially the same planarity of a peptide bond.

It will be appreciated that the peptide may conveniently be blocked at its N- or C-terminus so as to help reduce susceptibility to exoproteolytic digestion. Similarly, it will be appreciated that the peptide of the invention may be in salt form or may contain additional esters of -OH or -COOH groups or amides of -NH<sub>2</sub> groups. The peptides of the invention are defined in the claims, which would include any variants.

By a "variant" of the given amino acid sequence we mean that the side chains of one or two or three of the amino acid residues are altered (for example by replacing them with the side chain of another naturally occurring amino acid residue or some other side chain) such that the peptide is still able to bind to the HLA molecule in substantially the same way as a peptide consisting of the given amino acid sequence. For example, a peptide may be modified so that it at least maintains, if not improves, the ability to interact with and bind HLA-A\*1101 and so that it at least maintains, if not improves, the ability to generate activated CD8\* T cells which can recognise Hepatitis V virus. Typically, the amino acid alternatives are conservative in nature, such as from within the groups Gly, Ala; Ile, Leu, Val; Ser, Thr; Tyr, Phe, Trp; Glu, Asp; Gln, Asn, His, Met, Cys, Ser.

Peptides of at least 15 amino acids are preferred. Thus, the invention also includes peptides of 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 amino acids that contain an amino acid sequence selected from the group comprising: STLPETAVVRR, STLPETAVVR, 5 STLPETTVVRR, STLPETTVIRR, STPPETTVVRR, STLPETTVVGR and STIPETTVVRR. As noted above, the peptides of the invention are capable of binding HLA-A\*1101.

Those amino acid residues that are not essential to interact with the T cell receptor can be modified by replacement with another amino acid whose incorporation does not 10 substantially affect T cell reactivity and does not eliminate binding to the relevant HLA allele.

The peptides of the invention (and for use in the invention) are less than 5 000 and typically about 4 000 or 3 000 or 2 000. In terms of the number of amino acid residues, the peptides of the invention may have fewer than 30 or 20 or 19 or 18 or 17 or 16 or 15 or 14 or 13 or 15 12.

It will be appreciated from the following that in some applications the peptides of the invention may be used directly (i.e. they are not produced by expression of a polynucleotide in a patient's cell or in a cell given to a patient); in such applications it is preferred that the 20 peptide has fewer than 30 or 25 or 24 or 23 or 22 or 21 or 20 or 19 or 18 or 17 or 16 or 15 or 14 or 13 or 12 residues.

The peptides of the invention are able to bind to HLA-A\*1101. It is particularly preferred if the peptides bind selectively to HLA-A\*1101. 25

It is further preferred that the peptides of the invention are ones which can be used to generate peptide-specific CD8+ T cells which mediate specific killing of Hepatitis B virus.

The peptides of the invention (HBVcore169) are particularly useful in immunotherapeutic 30 methods to combat a Hepatitis B virus. In particular, the peptide in combination with the specific HLA molecule can be used to select and define appropriate T cells, and trace them once they are put into the patient. It is particularly preferred that in all the

immunotherapeutic methods of the invention that the patient to be treated is one who carries Class I HLA-A\*1101 (ie has a Class I HLA-A\*1101 -positive genotype), and has antigen presenting cells which express HLA-A\*1101.

- 5 The peptides of (and for use in) the invention are ones which bind HLA-A\*1101 and when so bound the HLA-A\*1101-peptide complex, when present on the surface of a suitable antigen-presenting cell, is capable of eliciting a T cell mediated immune response which mediates or helps to mediate the immune systems attack on Hepatitis B virus. In particular, the production of cytokines by a CD8+ T cell may mediate the attack on Hepatitis B virus.

10

Peptides (at least those containing peptide linkages between amino acid residues) may be synthesised using any method well known in the art, for example by the Fmoc-polyamide mode of solid-phase peptide synthesis as disclosed by Lu et al (1981) J. Org. Chem. 46, 3433 and references therein. Reagents for peptide synthesis are generally available from commercial providers of chemical and biological reagents. Purification may be effected by any one, or a combination of, techniques such as size exclusion chromatography, ion-exchange chromatography and (principally) reverse-phase high performance liquid chromatography. Analysis of peptides may be carried out using thin layer chromatography, reverse-phase high performance liquid chromatography, amino acid analysis after acid hydrolysis and by fast atom bombardment (FAB) mass spectrometric analysis.

15

20

25

In another aspect of the invention, there is provided a method for selecting Hepatitis B virus antigen specific T cells, the method comprising contacting a population of T cells with peptides according to the first aspect of the invention. Preferably, the method comprising contacting a population of T cells with a peptide or polypeptide of the invention presented in a HLA-A\*1101 molecule to which said peptide binds.

In an embodiment, the population of T cells are from an individual who has been exposed to a Hepatitis B virus.

30

It is appreciated that the peptides of the invention may be used to generate an expansion of T cells specific for Hepatitis B virus for patients who are HLA-A\*1101 positive and there are several ways in which the invention may be used.

- 5 The knowledge that the peptides of the invention bind the HLA-A\*1101 molecule and is only recognised by T cells in this context means that the creation of peptide-MHC multimers that will directly bind the T cells can occur. This reagent can be used to directly select the Hepatitis B virus specific cells from the bulk culture in order to infuse cells with a high purity into a patient.

10

The invention also includes use of the MHC multimer in conjunction with additional multimers of other specificities (to be defined) to fully characterise the product prior to infusion to ensure purity and safety of the product. Following infusion of Hepatitis B virus T cells to the patient the reconstitution of Hepatitis B virus specific immunity could be  
15 monitored in the patient using the peptide/HLA-A\*1101 multimer directly ex vivo.

An additional application of this technology would be that instead of relying on the innate antigen presentation ability of the patient's own cells to present the peptide of the invention in order to expand antigen presenting cells, an artificial antigen presenting cell  
20 could be used which consists of either cell lines deficient in all MHC but HLA-A\*1101 pulsed with the peptide or artificial antigen presenting cells which can be engineered to supply the co-stimulation required for T cell expansion along with the HLA-A\*1101 peptide combination in order to stimulate and expand the Hepatitis B virus specific T cells.

- 25 Suitable methods for selecting Hepatitis B virus-specific T cells include the use of ELISPOT analysis to confirm the responding T cells. Blood is obtained from HLA-A\*1101-typed donors or patients. Peripheral blood mononuclear cells (PBMC) are isolated via centrifugation in Biocoll Separating Solution (Biochrom, Berlin, Germany) and either used directly after preparation or cryopreserved for later use. Cells are cultured in RPMI 1640 with L-Glutamine  
30 (Invitrogen, Karlsruhe, Germany), supplemented with 10% heat-inactivated, pooled human serum and 100 U/ml Penicillin-Streptomycin (Invitrogen, Karlsruhe, Germany). Hepatitis B virus-specific T cell lines are generated by incubation of  $1 \times 10^7$  whole PBMC per well in 6-

well culture plates with the FHT peptide antigen for 7 days. Lymphocyte cultures are supplemented with 5 U/ml IL-2 (Proleukin, Chiron, Ratingen, Germany) every other day and culture medium replenished as needed. The T cell clones are generated by stimulating PBMCs repeatedly with 1 ua/ml FHT peptide once weekly for 4 weeks. Subsequently T cell clones are generated by limiting dilution in 96-well plates and expanded using the rapid expansion protocol as described by Beck et al. This small scale culture system may be scaled up and adapted to a "closed system" whereby clinical grade T cells suitable for infusion back into patients can be generated.

10 Suitable methods for selecting Hepatitis B virus-specific T cells include the use of a fluorescence-activated cell sorter (FACS). Following exposure of the donor or patient PBMCs to the peptides of the invention, the responding T cells are labeled based on activation markers or by behavioural characteristics. The labeling is achieved using an antibody specific for the activation marker or the secreted cytokine and such antibody is conjugated to a  
15 fluorochrome. The cells can then be separated and selected through a flow cytometer equipped for FACS analysis.

Alternatively, labeling of the responding T cells is based on the binding of an MHC multimer (HLA-A\*1101), which is conjugated to a fluorescent marker, to the specific TCR on the  
20 surface of the Hepatitis B virus-specific T cell.

Suitable methods for selecting T cells include the Cytokine Secretion Assay System which is manufactured by Miltenyi Biotec.

25 Methods for making and using peptide-loaded MHC multimers are described, for example, in Altman et al (1996) Science 274, 94-96; Kuabel et al (2002) Nature Medicine 8, 631-637; and Neudorfer et al (2007) J. Immunol. Methods 320, 119-131.

The purity of a T cell population may be assessed using the fluorescently labelled MHC  
30 multimer/peptide complex.

Suitable methods for selecting T cells also include the MHC Multimer System, available through Proimmune and Stage Pharmaceutical, which works by creating an artificial construction of HLA molecules which bind, in the present case, the peptides of the invention. These soluble, standalone HLA molecules may be constructed in a multimeric configuration so that a single multimer has 4-5 HLA molecules each loaded with a peptide of the invention. These multimers may also be attached to a magnetic bead as above. The multimers are released into a blood sample, and the HLA:peptide construct will bind with T cell receptors that recognise the peptide and hence will label the T cells which will recognise and mount an immune response against Hepatitis B virus. The cell sample is passed through a magnetic column, and the labelled cells are retained and then released.

It will be appreciated from the above that the invention may include a complex comprising a HLA-A\*1101 molecule bound to a peptide according to the first aspect of the invention. Conveniently, the complex is a soluble complex and is not cell-bound.

Preferably, the peptide in the complex is any one of the peptides of the invention, but may be any other peptide of the invention that will form a complex, and be useful in eliciting an anti-Hepatitis B virus T cell response. As is plain, the complex may be used for isolating Hepatitis B virus -specific T cells. The complex may also be used to identify an Hepatitis B virus-specific T cell in a sample.

Typically, the peptide is presented on a dendritic cell or monocyte which present the antigen to T cells. The T cells then secrete cytokines which in turn activate monocytes and neutrophils for enhanced killing of Hepatitis B virus.

In an aspect of the invention, there is provided A T cell expressing a T cell receptor (TCR) molecule, wherein the TCR molecule comprises an amino acid sequence selected from the group comprising: CASGDSNSPLHF, CASSGGQIVYEQYF, CSARGGRGGDYTF and CASSQDWTEAFF.

The invention also includes T cells, preferably CD8+ T cells, which have been transfected with a polynucleotide or expression vector which expressed the above mentioned TCR or

functionally equivalent molecule. The T cells may be obtained from the patient or, in the case of an allogeneic HSCT patient, from a closely matched donor with respect to HLA type.

More particularly, the T cell of the invention is for use in inducing antiviral T cell activity against a Hepatitis B virus core169 epitope, wherein a polynucleotide or an expression vector of the invention has been introduced into the T cell, preferably patient-derived T cell, so that the T cell expresses the encoded TCR molecule.

As well as the TCR molecule, functionally equivalent molecules to the TCR are included in the invention. These include any molecule which is functionally equivalent to a TCR which can perform the same function as a TCR. In particular, such molecules include genetically engineered three-domain single-chain TCRs as made by the method described by Chung et al (1994) Proc. Natl. Acad. Sci. USA 91, 12654-12658, and referred to above.

Typically, the TCR or a functionally equivalent molecule to the TCR, recognises a human Class I HLA molecule expressed on the surface of an antigen-presenting cell and loaded with a peptide according to the first aspect of the invention.

In various embodiments of the invention, the Hepatitis B virus-specific T cell may be isolated for further use. With some techniques it is possible to isolate a sufficient number of specific T cells for therapeutic use directly, but it may be necessary to expand or clone the T cells to produce a sufficient number. For adoptive immunotherapy it is generally preferred to use a technique which allows for the isolation of sufficient numbers of cells directly since this can be achieved within a day (whereas cell expansion may take several weeks).

A suitable procedure for identifying pathogen-specific donor clones is described in Perruccio et al (2005) Blood 106, 4397-4406.

The Hepatitis B virus-specific T cells which are directed against the peptides of the invention are useful in therapy.

It will be appreciated that the skilled person can readily design and synthesise TCRs according to the invention using either or any nomenclature systems provided that the framework region (ie region not replaced by the CDRs) is compatible with the CDRs as is well known in the art.

5

By "TCR molecule" we include any molecule which contains the given CDRs and also contains FRs suitably situated within the molecule so that the TCR forms a recognition site (binding site) which is able to bind to HLA-A\*11:01 presenting any one of the following peptides: STLPETAVVRR, STLPETAVVR, STLPETTVVRR, STLPETTVIRR, STPPETTVVRR, STLPETTVVGR and STIPETTVVRR.

10

It is particularly preferred if the TCR molecules contain the precise amino acid sequences CASGDSNSPLHF, CASSGGQIVYEQYF, CSARGGRGGDYTF and CASSQDWTEAFF. Where a variant to this precise sequence is present, it preferably varies by one or two or three (preferably one or two) amino acids. Typically, in these variants, the amino acids which are replaced are replaced with conservative amino acids. By conservative amino acids we include the groupings: G, A; S, A, T; F, Y, W; D, E; N, Q; and I, L, V.

15

In various embodiments, the amino acid encodes for a TCR beta chain portion, particularly CDR1, CDR1 and CD3 of the beta chain portion.

20

In another aspect of the invention, there is provided at least one polynucleotide encoding the TCR molecule as defined above. In an embodiment, the polynucleotide comprises sequences SEQ ID NOS 1 to 16 set out in the table below.

25

| SEQ ID NO. | TCR molecule amino acid sequence | Polynucleotide sequence                                                                         |
|------------|----------------------------------|-------------------------------------------------------------------------------------------------|
| 1          | CASGDSNSPLHF                     | CTAAACCTGAGCTCTCTGGAGCTGGGGGACTCAGCTTTGTAT<br>TTCTGTGCCAGCGGGGATTCCAATTCACCCCTCCACTTTGGGA<br>AC |
| 2          | CASGDSNSPLHF                     | ACTAAACTGAGCTCTCTGGAGCTGGGGGACTCAGCTTTGTAT<br>TTCTGTGCCAGCGGGGATTCCAATTCACCCCTCCACTTTGGGA       |

|    |                |                                                                                                  |
|----|----------------|--------------------------------------------------------------------------------------------------|
|    |                | AC                                                                                               |
| 3  | CASGDSNSPLHF   | ACTAAACCTGAGCTCTCTGGAGCTGGGGACTCAGCTTTGTAT<br>TTCTGTGCCAGCGGGGATTCCAATTCACCCCTCCACTTTGGGA<br>AC  |
| 4  | CASGDSNSPLHF   | ACTAACCTGAGCTCTCTGGAGCTGGGGGACTCAGCTTTGTAT<br>TTCTGTGCCAGCGGGGATTCCAATTCACCCCTCCACTTTGGGA<br>AC  |
| 5  | CASGDSNSPLHF   | TAAACCTGAGCTCTCTGGAGCTGGGGGACTCAGCTTTGTAT<br>TTCTGTGCCAGCGGGGATTCCAATTCACCCCTCCACTTTGGGA<br>AC   |
| 6  | CASGDSNSPLHF   | CTAAACCTGAGCTCTCTGGAGCTGGGGGACTCAGCTTTGTAT<br>TTCTGTGCCAGCGGGGATTCCAATTCACCCCTCCACTTTGGGA<br>AC  |
| 7  | CASGDSNSPLHF   | CTAAACCTGAGCTCTCTGGAGCTGGGGGACTCAGCTTTGTAT<br>TTCTGTGCCAGCGGGGATTCCAATTCACCCCTCCACTTTGGGA<br>AC  |
| 8  | CASSGGQIVYEQYF | GTGAACGCCTTGTTGCTGGGGGACTCGGCCCTGTATCTCTGT<br>GCCAGCAGCGGGGGACAGATTGTATACGAGCAGTACTTCGG<br>GCCG  |
| 9  | CASSGGQIVYEQYF | TGTGAACGCCTTGTTGCTGGGGGACTCGGCCCTGTATCTCTGT<br>GCCAGCAGCGGGGGACAGATTGTATACGAGCAGTACTTCGG<br>GCCG |
| 10 | CASSGGQIVYEQYF | GTGAACGCCTTGGAGCTGGACGACTCGGCCCTGTATCTCTGT<br>GCCAGCAGCGGGGGACAGATTGTATACGAGCAGTACTTCGG<br>GCCG  |
| 11 | CASSGGQIVYEQYF | GTGAACGCCTTGTTGCTGGGGGACTCGGCCCTGTATCTCTGT<br>GCCAGCAGCGGGGGACAGATTGTATACGAGCAGTACTTCGG<br>GCCG  |
| 12 | CSARGGRGGDYTF  | ACAGTGACCAAGTGGCCATCCTGAAGACAGCAGCTTCTACATC<br>TGCAGTGCAAGGGGAGGAAGGGGCGGAGACTACACCTTCG<br>GTTCG |

|    |               |                                                                                                  |
|----|---------------|--------------------------------------------------------------------------------------------------|
|    |               |                                                                                                  |
| 13 | CSARGGRGGDYTF | ACAGTGACCAAGTGGCCATCCTGAAGACAGCAGCTTCTACATC<br>TGCAGTGCAAGGGGAGGAAGGGGCGGAGACTACACCTTCG<br>GTTCG |
| 14 | CSARGGRGGDYTF | ACAGTGACCAAGTGGCCATCCTGAAGACAGCAGCTTCTACATC<br>TGCAGTGCGAGGGGAGGAAGGGGCGGAGACTACACCTTCG<br>GTTCG |
| 15 | CASSQDWTEAFF  | CTGAAGGTGCAGCCTGCAGAACTGGAGGATTCTGGAGTTTA<br>TTTCTGTGCCAGCAGCCAAGACTGGACTGAAGCTTTCTTTGG<br>ACAA  |
| 16 | CASSQDWTEAFF  | CTGAAGGTGCAGCCTGCAGAACTGGAGGATTCTGGAGTTTA<br>TTTCTGTGCCAGCAGCCAAGATTGGACTGAAGCTTTCTTTGG<br>ACAA  |

The polynucleotide may be DNA or RNA, and it may or may not contain introns. Typically, the polynucleotide does not contain introns within the region that codes for the polypeptide of interest. It will be appreciated that different polynucleotides may encode the same polypeptide because of the degeneracy of the genetic code.

It will be appreciated that the polynucleotide is typically used in an HLA-A\*1101-specific context.

The invention also provides an expression vector that contains the polynucleotide encoding the peptides and TCR molecule of the invention. Such expression vectors, when present in a suitable host cell, allow for the expression of the polypeptide(s) of interest. Preferably, the expression vector is an expression vector capable of expressing a polypeptide in a mammalian cell. More preferably, the expression vector is one which is able to express a polypeptide in a T cell, such as a human CTL. Typically, the expression vectors contain a promoter which is active in particular cell types, and which may be controllable (eg inducible).

It will be appreciated that the expression vector is typically used in an HLA-A\*1101-specific context. The vector is suitably a retroviral vector which is capable of transfection into a mammalian host cell such as a human T cell. Typically, the vector is a lentiviral vector.

- 5 Other suitable expression vectors include viral based vectors such as retroviral or adenoviral or vaccinia virus vectors or lentiviral vectors or replication deficient MV vectors. Suitable general cloning vectors include plasmids, bacteriophages (including  $\lambda$  and filamentous bacteriophage), phagemids and cosmids.
- 10 Methods for manipulating, changing and cloning nucleic acid molecules are well known in the art, for example Sambrook J and Russell, DW, Molecular Cloning, A Laboratory Manual, 3rd Edition, 2001, Cold Spring Harbor Laboratory Press describes such techniques including PCR methods.
- 15 A further aspect of the invention provides a host cell comprising a polynucleotide encoding the peptides and TCR molecules of the invention or a vector of the invention. The host cell may contain a polynucleotide or vector which encodes only the alpha chain portion or only the beta chain portion. However, if the host cell is to produce a TCR molecule of the invention, it contains one or more polynucleotides or vectors which encode both the alpha
- 20 chain portion and the beta chain portion.

In various embodiments, the host cell is a T cell derived from an individual.

- The host cell may be any cell such as a bacterial cell (e.g. Escherichia coli, Bacillus subtilis and Salmonella typhimurium), yeast cell (e.g. Saccharomyces cerevisiae and Schizosaccharomyces pombe), insect cell, plant cell or mammalian cell (e.g. CHO cells, COS
- 25 cells and other mammalian cells such as antigen presenting cells), and methods of introducing polynucleotides into such cells are well known in the art. Typically, bacterial cells, such as Escherichia coli cells are used for general propagation and manipulation of the
- 30 polynucleotides and vectors of the invention. Other host cells may be used to express the TCR molecules of the invention and, in particular, the cell may be a mammalian cell such as a human cell. As described below in relation to the therapeutic methods using

the TCR molecules of the invention, it is particularly desirable if the host cell is a T cell such as (and preferably) a T cell derived from a patient to be treated, typically a patient with a WT1-expressing malignancy.

- 5 Typically, a retroviral vector (or, as the case may be vectors) encoding the TCR molecule of the invention is used based on its ability to infect mature human CD4+ or CD8+ T lymphocytes and to mediate gene expression: the retroviral vector system Kat is one preferred possibility (see Finer et al (1994) Blood 83, 43). High titre amphotrophic retrovirus are used to infect purified CD8+ T lymphocytes isolated from the peripheral blood of tumour  
10 patients following a protocol published by Roberts et al (1994) Blood 84, 2878-2889, incorporated herein by reference. Anti-CD3 antibodies are used to trigger proliferation T cells, which facilitates retroviral integration and stable expression of single chain TCRs. A combination of anti-CD3 and anti-CD8 antibodies may be more effective than anti-CD3 antibodies alone. Other suitable systems for introducing genes into CTL are described in  
15 Moritz et al (1994) Proc.Natl. Acad. Sci. USA 91, 4318-4322, incorporated herein by reference. Eshhar et al (1993) Proc.Natl. Acad. Sci. USA 90, 720-724 and Hwu et al (1993) J. Exp. Med. 178, 361-366 also describe the transfection of CTL. The commercially available Nucleofactor system, provided by AMAXA, Germany may be used to transfect T cells. Retroviral transduction of human CD8+ T cells is described in Stanislawski (2001) Nat.  
20 Immunol. 2, 962. Methods of cloning and genetic manipulation are well known in the art and are described in detail in standard manuals such as Sambrook & Russell (2001) Molecular Cloning, a laboratory manual, Cold Spring Harbor Press, Cold Spring Harbor, NY, USA.
- 25 In another aspect of the invention, there is provided a T cell according to an earlier aspect of the invention for use in inducing antiviral T cell activity against a Hepatitis B virus, and use of said T cell in the preparation of a medicament for inducing antiviral T cell activity against a Hepatitis B virus.
- 30 In an embodiment, the Hepatitis B virus expresses an epitope comprising an amino acid sequence selected from the group comprising: STLPETAVVRR, STLPETAVVR, STLPETTVVRR, STLPETTVIRR, STPPETTVVRR, STLPETTVVGR and STIPETTVVRR.

The phrase " inducing antiviral T cell activity against a Hepatitis B virus " is understood to mean that the antigen binding site of the present invention induces or enhances any one or more activities of the T cell, including but not limited to stimulating proliferation, cytotoxicity or maturation of NK cells, stimulating proliferation or differentiation of B cells and T cells; stimulating antibody production and affinity maturation in B cells; stimulating cytotoxicity of CD8+ T cells; stimulating interferon gamma production in T-cells and NK-cells; inhibiting dendritic cell (DC) activation and maturation; inhibiting release of inflammatory mediators from mast cells, enhancing phagocytosis of macrophages, inhibiting generation or survival of TReg cells, and stimulating the proliferation of bone marrow progenitor cells.

The peptides, TCR molecules, polypeptides, polynucleotides, expression vectors and T cells of the invention may be packaged and presented for use as a medicament. In particular, they are of use in treating or combating a Hepatitis B virus infection. By " Hepatitis B virus infection " we include treating patients who have Hepatitis B virus infection. We also include administering the peptides, TCR molecules, polypeptides, polynucleotides, expression vectors and T cells of the invention (either alone or in combination with, or present in or on, a suitable HLA matched antigen presenting cell such as a dendritic cell or B cell or monocytes or a synthetic APC) to not only patients who have a Hepatitis B virus infection, but also to those at risk of Hepatitis B virus infection.

Patients at risk of Hepatitis B virus infection include those who are immunocompromised or immunodepleted such as those undergoing allogeneic HSCT, organ transplant patients, autoimmune patients receiving immunosuppressive drugs, patients with genetic immune disorders, AIDS patients, or patients undergoing chemotherapy for cancer or leukaemia patients. Thus, it will be appreciated that "combating" includes preventing (or helping to prevent) Hepatitis B virus infection and treating a patient prophylactically.

In another aspect of the invention, there is provided a method of treating a Hepatitis B virus infection in an individual, the method comprising administering to the individual an effective amount of a peptide according to the first aspect of the invention or a

polynucleotide according to various aspects of the invention described above or a T cell according to an aspect of the invention described above.

The term "effective amount" or "therapeutically effective amount" means a dosage sufficient to treat, inhibit, or alleviate one or more symptoms of a disease state being treated or to otherwise provide a desired pharmacologic and/or physiologic effect, especially enhancing T cell response to a selected antigen. The precise dosage will vary according to a variety of factors such as subject-dependent variables (e.g., age, immune system health, etc.), the disease, and the treatment being administered.

In another aspect of the invention, there is provided a pharmaceutical composition comprising a peptide according to the first aspect of the invention or a polynucleotide according an aspect of the invention or a T cell according an aspect of the invention and a pharmaceutically acceptable carrier.

A pharmaceutically acceptable carrier refers, generally, to materials that are suitable for administration to a subject wherein the carrier is not biologically harmful, or otherwise, causes undesirable effects. Such carriers are typically inert ingredients of a medicament. Typically a carrier is administered to a subject along with an active ingredient without causing any undesirable biological effects or interacting in a deleterious manner with any of the other components of a pharmaceutical composition in which it is contained. Suitable pharmaceutical carriers are described in Martin, Remington's Pharmaceutical Sciences, 18th Ed., Mack Publishing Co., Easton, Pa., (1990), incorporated by reference herein in its entirety.

In a more specific form of the disclosure there are provided pharmaceutical compositions comprising therapeutically effective amounts of the peptide, polynucleotide and T cell together with pharmaceutically acceptable diluents, preservatives, solubilizers, emulsifiers, adjuvants and/or carriers. Such compositions include diluents of various buffer content (e.g., phosphate, Tris-HCl, acetate), pH and ionic strength and additives such as detergents and solubilizing agents (e.g., Tween 80, Polysorbate 80), anti-oxidants (e.g., ascorbic acid, sodium metabisulfite), preservatives (e.g., Thimersol, benzyl alcohol) and bulking substances (e.g., lactose, mannitol). The material may be incorporated into particulate preparations of

polymeric compounds such as, for example and without limitation, polylactic acid or polyglycolic acid, or into liposomes. Hylauronic acid may also be used. Such compositions may influence the physical state, stability, rate of in vivo release, and rate of in vivo clearance of the disclosed compositions. The compositions may be prepared in liquid form, or may be in dried powder, such as lyophilized form.

It will be appreciated that pharmaceutical compositions provided according to the disclosure may be administered by any means known in the art. Preferably, the pharmaceutical compositions for administration are administered by injection, orally, or by the pulmonary, or nasal route. The antisense polynucleotides are, in various embodiments, delivered by intravenous, intra-arterial, intraperitoneal, intramuscular, or subcutaneous routes of administration. In various embodiments, the composition is suitable for parenteral administration either naked or complexed with a delivery agent to a patient. The carrier may be selected from the group comprising of a nanoparticle, such as a polymeric nanoparticle; a liposome, such as pH-sensitive liposome, an antibody conjugated liposome; a viral vector, a cationic lipid, a polymer, a UsnRNA, such as U7 snRNA and a cell penetrating peptide.

The compositions or molecules of the invention may be administered orally, or rectal, or transmucosal, or intestinal, or intramuscular, or subcutaneous, or intramedullary, or intrathecal, or direct intraventricular, or intravenous, or intravitreal, or intraperitoneal, or intranasal, or intraocular.

The compositions or molecules of the invention encompass any pharmaceutically acceptable salts, esters, or salts of such esters, or any other compound which, upon administration to an animal including a human, is capable of providing (directly or indirectly) the biologically active metabolite or residue thereof. Accordingly, for example, the disclosure is also drawn to prodrugs and pharmaceutically acceptable salts of the compounds of the invention, pharmaceutically acceptable salts of such pro-drugs, and other bioequivalents.

The term "pharmaceutically acceptable salts" refers to physiologically and pharmaceutically acceptable salts of the compounds of the invention: i.e., salts that retain the desired biological activity of the parent compound and do not impart undesired toxicological effects thereto.

5

For polynucleotides, preferred examples of pharmaceutically acceptable salts include, but are not limited to, (a) salts formed with cations such as sodium, potassium, ammonium, magnesium, calcium, polyamines such as spermine and spermidine; (b) acid addition salts formed with inorganic acids, for example hydrochloric acid, hydrobromic acid, sulfuric acid, 10 phosphoric acid, nitric acid; (c) salts formed with organic acids such as, for example, acetic acid, oxalic acid, tartaric acid, succinic acid, maleic acid, fumaric acid, gluconic acid, citric acid, malic acid, ascorbic acid, benzoic acid, tannic acid, palmitic acid, alginic acid, polyglutamic acid, naphthalenesulfonic acid, methanesulfonic acid, p-toluenesulfonic acid, naphthalenedisulfonic acid, polygalacturonic acid; and (d) salts formed from elemental 15 anions such as chlorine, bromine, and iodine. The pharmaceutical compositions of the disclosure may be administered in a number of ways depending upon whether local or systemic treatment is desired and upon the area to be treated. Administration may be topical (including ophthalmic and to mucous membranes including rectal delivery), pulmonary, e.g., by inhalation of powders or aerosols, (including by nebulizer, intratracheal, 20 intranasal, epidermal and transdermal), oral or parenteral. Parenteral administration includes intravenous, intra-arterial, subcutaneous, intraperitoneal or intramuscular injection or infusion; or intracranial, e.g., intrathecal or intraventricular, administration. Polynucleotides with at least one 2'-O-methoxyethyl modification are believed to be particularly useful for oral administration.

25

The pharmaceutical formulations of the disclosure, which may conveniently be presented in unit dosage form, may be prepared according to conventional techniques well known in the pharmaceutical industry. Such techniques include the step of bringing into association the active ingredients with the pharmaceutical carrier(s) or excipient(s). In general the 30 formulations are prepared by uniformly bringing into association the active ingredients with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

Combination therapy with an additional therapeutic agent is also contemplated by the disclosure. Examples of therapeutic agents that may be delivered concomitantly with a composition of the disclosure include, without limitation, a glucocorticoid steroid (for  
5 example and without limitation, prednisone and deflazacort), an angiotensin converting enzyme inhibitor, a beta adrenergic receptor blocker, an anti-fibrotic agent and a combination thereof.

In some embodiments, the present invention may be used in gene therapy such, e.g. using a  
10 vector (e.g., an expression vector) comprising a polynucleotide of the invention to direct expression of the polynucleotide in a suitable host cell. Such vectors are useful, e.g., for amplifying the polynucleotides in host cells to create useful quantities thereof, and for expressing proteins using recombinant techniques. In some embodiments, the vector is an expression vector wherein a polynucleotide of the invention is operatively linked to a  
15 polynucleotide comprising an expression control sequence.

The peptides of the invention, alone or in combination with antigens from other pathogen, may be used to activate immune cells within a blood or tissue sample or a cellular derivative thereof obtained from the patient or a donor without significant further selection or  
20 purification of cell types (an "Unselected Cell Formulation") with a view to infusing the Unselected Cell Formulation in a patient in order to treat or prevent infection by Hepatitis B virus whether on a targeted basis or as one of several pathogens which may cause infection in a patient.

25 The methods of combating or treating Hepatitis B virus infection, and the pharmaceutical compositions and medicaments, of the invention may be combined with other anti-Hepatitis B virus treatments.

In another aspect of the invention, there is provided vaccine against Hepatitis B virus  
30 infection comprising a peptide according to the first aspect of the invention or a polynucleotide or a T cell according to various aspects of the invention described above, and may be packaged and presented for use as a medicament.

In another aspect of the invention, there is provided a method of combating a Hepatitis B virus infection in a patient which carries HLA A\*1101, the method comprising: (a) obtaining T cells from the patient; (b) introducing into the T cells a polynucleotide encoding a TCR molecule according to an aspect of the invention; (c) introducing the T cells produced in step (b) into the patient.

The transfected T cells are able to help fight off the Hepatitis B virus. Preferably, the patients to be treated carry Class I HLA A\*1101.

In various embodiments, the polynucleotide(s) is/are transfected or introduced to the T cells by electroporation. Other suitable systems for introducing genes into T cells are described in Moritz et al (1994) Proc. Natl. Acad. Sci. USA 91, 4318-4322. Eshhar et al (1993) Proc. Natl. Acad. Sci. USA 90, 720-724 and Hwu et al (1993) J. Exp. Med. 178, 361-366 also describe the transfection of T cells.

Methods for introducing a nucleic acid into the T cell, which are well known and routinely practiced in the art, include transformation, transfection, electroporation, nuclear injection, or fusion with carriers such as liposomes, micelles, ghost cells, and protoplasts. The host T-cell may be isolated and/or purified. The T-cell also may be a cell transformed in vivo to cause transient or permanent expression of the polypeptide in vivo. The T-cell may also be an isolated cell transformed ex vivo and introduced post-transformation, e.g., to produce the polypeptide in vivo for therapeutic purposes.

The polynucleotides of the present invention (particular SEQ ID Nos. 1 to 16) may be introduced into a T cell by methods of transfection well known in the art. These methods include sonophoresis, electric pulsing, electroporation, osmotic shock, calcium phosphate precipitation, and DEAE dextran transfection, lipid mediated delivery, passive delivery etc. The language "transfecting T cells" is intended to include any means by which a nucleic acid molecule can be introduced into a T cell. The term "transfection" encompasses a variety of techniques useful for introduction of nucleic acids into mammalian cells including electroporation, calcium-phosphate precipitation, DEAE-dextran treatment, lipofection,

microinjection, and viral infection. Suitable methods for transfecting mammalian cells can be found in Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory press (1989)) and other laboratory textbooks.

- 5 The polynucleotides may also be introduced into a T cell using a viral vector. Such viral vectors include, for example, recombinant retroviruses, adenovirus, adeno-associated virus, and herpes simplex virus-1. Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery system of choice for the transfer of exogenous genes in vivo, particularly into humans. Alternatively they can be used for
- 10 introducing exogenous genes ex vivo into T cells. These vectors provide efficient delivery of genes into T cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host cell.

- Another viral gene delivery system useful in the present invention utilizes adenovirus-
- 15 derived vectors. The genome of an adenovirus can be manipulated such that it encodes and expresses a gene product of interest but is inactivated in terms of its ability to replicate in a normal lytic viral life cycle. Yet another viral vector system useful for delivery of a nucleic acid molecule comprising a gene of interest is the adeno-associated virus.

- 20 The polynucleotides may be carried by and delivered into a T cell by a cell-delivery vehicle. Such vehicles include, for example, cationic liposomes (Lipofectin™) or derivatized (e.g. antibody conjugated) polylysine conjugates, gramicidin S, artificial viral envelopes. These vehicles can deliver a nucleic acid that is incorporated into a plasmid, vector, or viral DNA. In a specific embodiment, efficient introduction of the nucleic acid molecule in primary T
- 25 lymphocytes is obtained by transfecting the primary T lymphocytes with adeno-associated virus plasmid DNA complexed to cationic liposomes, as described in Philip, R. et al. (1994) Mol. Cell. Biol. 14, 2411.

- In another embodiment of the invention, the polynucleotides may be delivered in the form
- 30 of a soluble molecular complex. The complex contains the nucleic acid releasably bound to a carrier comprised of a nucleic acid binding agent and a cell-specific binding agent which

binds to a surface molecule of the specific T cell and is of a size that can be subsequently internalized by the cell.

In another embodiment of the invention the polynucleotides is introduced into T cells by particle bombardment.

In various embodiments, the polynucleotides may be passively delivered (i.e., deliver without additional transfection reagents) to the T cells, particularly during T cell expansion. Expansion as used herein includes the production of progeny cells by a transfected neural stem cell in containers and under conditions well know in the art. Expansion may occur in the presence of suitable media and cellular growth factors. The polynucleotides may be passively delivered to the T cells in culture (e.g., in culture plates, culture dishes, multiwell plates etc without limitation) under reduced serum conditions, including under 0% serum conditions. Such conditions include cells cultured in standard, art-tested reduced-serum media that are commercially available from numerous companies including Invitrogen, and HyClone. In one example, cells are first plated in serum medium, then the serum medium is replaced with reduced serum medium comprising a tripartite oligonucleotide complex of the disclosure for 24 hours, then the reduced serum medium is replaced with serum medium.

In various embodiments, the transfection reagent may be selected from the group consisting of polymers, lipids, lipid-polymers and/or their combinations and/or their derivatives containing a cell-targeting or an intracellular targeting moiety and/or a membrane-destabilizing component and one or more delivery enhancers.

In another aspect of the invention, the peptides according to the first aspect of the invention may be used to create a monoclonal or polyclonal antibody either on a patient-specific basis or batch manufacturing basis wherein the antibody is used to prevent or treat infection by Hepatitis B virus or to induce a primary or secondary humoral or cellular immune response to Hepatitis B virus in a patient. The antibody will include idiomatic derivations of antibodies specific for the peptides described above. Preferably, the antibody recognises the peptides of the invention when presented by a Class I HLA-A\*1101 molecule.

In order that the present invention may be fully understood and readily put into practical effect, there shall now be described by way of non-limitative examples only preferred embodiments of the present invention, the description being with reference to the accompanying illustrative figures.

Fig. 1. Comprehensive epitope mapping against HBV using highly multiplexed combinatorial pMHC tetramer strategy. (A) The experimental workflows. The 562-plex pMHC tetramer library was generated from the deep sequencing of virus and epitope prediction. The library included 484 putative A\*11:01-restricted HBV peptides and 78 known control peptides derived from other common virus or self-antigens. 1001 unique combinations of quadruple SAv-metal codings were used to code the entire library. A self-validated tetramer deconvolution algorithm automatically identified the signals on patient's T cells with statistical measurements. Validated antigen-specific CD8+ T cells targeting four viral epitopes were shown. (B) Mean frequency of HBV-specific CD8+ T cells from all patients tested across four different viral proteins. Plot only shows the detectable epitopes. Numbers at the bottom indicate the numbers of epitopes detected/screened for each viral protein. (C) Epitopes nomenclature and annotation used in this report are shown. \* indicates a peptide cluster that contained more than one peptide (table S1). Peptide sequences in bold faces are previous unpublished sequences based on Immune Epitope Database (IEDB). (D) The frequencies of four antigen-specific CD8+ T cells across various patient groups as color-coded. (E) The expression of cellular markers on four HBV-specific CD8+ T cells are shown in heatmaps. Boxes highlight the discriminative markers for each patient group.

Fig. 2. Multifactorial memory atlas of HBVpol387 and HBVcore169-specific CD8+ T cells linked to HBV clinical stages. (A) Unsupervised Phenograph clustering of cellular subsets on all detected antigen-specific CD8+ T cells across patient groups. n=20, 4 patients per group. Nineteen cellular clusters objectively identified by Phenograph were color-coded as indicated, and the expression levels of probed cellular proteins are shown. (B) Visualization of the Phenograph clustering of nine major cellular clusters of HBVpol387-specific CD8+ T cells is shown. The proportion of cellular clusters within HBVpol387-specific CD8+ T cells in

individuals across various patient groups are shown. (C) The same analytical strategy for HBVcore169-specific CD8+ T cells is shown. (D) Bar graph indicates the discrepancy of T cell memory-associated markers (CD27, CD28, CD45RO, CD127 and CXCR3), inhibitory receptors (PD-1 and TIGIT) and CD57 expressed on HBVcore169-specific CD8+ T cells. Error bars are median and range, and values from individuals were imposed. (E) Representative contour plots show the expression level of markers on HBVcore169-specific CD8+ T cells between patient groups. Patients were color coded as indicated. (F) Logistic regression (Upper panel) of 8 phenotypic markers that showed significant difference between patient groups were stacked against pseudotime imputed using Scorpius. Logistic regression (black solid line, lower panel) was used to visualize the trend of these 8 cellular markers. Dots are individuals color-coded by clinical stages, and showed the expression levels of these markers on HBVcore169-specific CD8+ T cells.

Fig. 3. Unsupervised analyses uncovered the complex model of inhibitory receptors (exhaustion markers) in CHB. (A) One-SENSE objectively related three different T cell categories (Differentiation+TNFR, Inhibitory and Trafficking) in 2D plots with visualization of the expression levels of cellular proteins. Dots are selected virus-specific CD8+ T cells as color-coded. Boxes annotated the epitopes who were enriched in the given regions. (B) The average fractions of the numbers of co-expressed inhibitory receptors on four HBV-specific CD8+ T cells across patient groups. Plots were from a representative experiment with all nine inhibitory receptors. (C) The average co-expressed inhibitory receptors on four on four HBV-specific CD8+ T cells across patient groups. Plots were comprised of three experiments with simultaneously measurements of eight inhibitory receptors (without TIGIT). Each dot is an individual.

Fig. 4. Nonlinear correlations of multi-functionalities and inhibitory receptors by One-SENSE (A) Patient's PBMCs were stimulated with correspondent viral peptides for 10 days in vitro culture to measure the functional capacity. Categorical (Function, Inhibitory and Differentiation+TNFR) analysis of One-SENSE revealed the diverse multi-functional virus-specific CD8+ T cells subsets and their corresponded co-expressions of inhibitory receptors. Dots are different virus-specific CD8+ T cells as annotated. Five different major functional subsets were labeled based on the aligned heatplots and color-coded as indicated. (B) The

expression levels of T cells functions, inhibitory receptors and TNFR costimulatory receptors were compared between these five functional subsets. (C) Bar graphs showed the proportion of each functional subset in HBVpol387 and HBVcore169-specific CD8+ T cells across patient groups. n=5 per group except for IT=4.

5

Fig. 5. Unsupervised quantifications of HBV-specific TCR were associated with disease stages in an epitope-dependent manner. (A) TCRdist measurements of epitope-specific TCRs were clustered by unsupervised Phenograph analysis and then projected by t-SNE. Each dot is one TCR clone. Twenty-eight TCR-sequence clusters on t-SNE map were labeled. (B) The sequence motifs (dashed boxes labeled with size) of representative TCR-sequence clusters were shown. Average-linkage dendrogram for each TCR in the given cluster was presented and color-coded by generation probability. TCR logos display the frequency of V and J segments with CDR3 sequence in the middle. Bottom bars are source region as indicated. Light grey is V-region. Red is N-insertion. Black is D for diversity. Dark grey is J-region. (C) The percentages of the receptor in total epitope-specific TCRs for twenty-eight TCR-sequence clusters were shown. (D) The proportion of TCR cluster 27 (C27) and 15 (C15) in four different epitope-specific TCRs. (E) TCRdiv diversity measures for each epitope-specific TCR across patient groups. (F) Stacked bar charts show the Top eleven TCR clones in individual patient. The frequencies and sequences of public TCR clones were presented. (G) 3D PCA projection delineated patient's clinical stage using the epitope-specific TCR repertoires, tetramer response and cellular profiles from the same individuals. (H) Correlation between the frequency and TCRdiv diversity measures of HBVcore169-specific CD8+ T cells.

10

15

20

25

30

Fig. 6 The phenotypic dynamic and the machine learning-aided modeling of HBVcore169-specific CD8+ T cells. (A) A total of 14 patients (n=8 for HBeAg-, and n=6 for HBeAg+) were included in the longitudinal cohort. The average frequency of all detectable HBVcore169-specific CD8+ T cells across different time points was shown. Each dot is one patient who had detectable HBVcore169-specific CD8+ T cells. (B) Dynamics of HBVcore169-specific CD8+ T cells in two representative patients. (C) The phenotypic dynamic of HBVcore169-specific CD8+ T cells was shown using One-SENSE. Numbers are frequencies and boxes are annotated as indicated. (D) The fractions of memory (blue boxes) and terminal effector (red

boxes) cells in individual patients across longitudinal time points. (E) Plot showed the changes of selective cellular markers expression on HBVcore169-specific CD8+ T cells across patient's longitudinal time points (early and late, thick stacked bars in D). Two time points (early and late) were picked to roughly match the time points between patients based on the drug intervention. "Early" are pre-treatment time points besides one patient (HBeAg+04, whose earliest time point was 3 months post-treatment), and "Late" are roughly 30 months post-treatment. Plots showed the patients who had detectable HBVcore169-specific CD8+ T cells in both early and late time points. Statistical analysis was used to compare the cellular marker expressions between two time points (early and late, solid line), or patient groups (HBeAg+ and HBeAg-, dash line). (F) Logistic model (dashed grey line) of cellular markers expression (dependent variable) against SVM-predicted pseudotime (independent variable). Dots represent the expression levels of cellular markers on HBVcore169-specific CD8+ T cells across different patient's longitudinal time points. (G) Statistical analysis of SVM-predicted pseudotime during the progression of patient's longitudinal time points. A non-parametric paired t-test was used.

Fig. 7 Comprehensive epitope mapping strategy and experimental workflow.

(A) HBV was deep-sequenced using next-generation sequencing (NGS). NetMHC (v3.4) was used to predict HLA-A\*11:01-restricted epitopes based on the consensus sequences. (B) 562 peptides were clustered based on the sequence homology and were further given unique combinations of four SAv-metal codes. (C) These unique combinations of four SAv-metal mixtures for two different coding configurations were prepared by automatic liquid handling robot. The quadruple SAv-metal coded pMHC tetramer library (for two coding configurations) was used to stain on patient's PBMC. (D) Tetramer positive cells were determined by an automatic tetramer deconvolution algorithm and the tetramer signals between the two coding configurations were calculated for their correspondence.

Fig. 8 The quality and detection of antigen-specific CD8+ T cells using highly multiplexed combinatorial pMHC tetramer staining and mass cytometry.

(A) The staining quality of quadruple SAv-metal coded pMHC tetramers using fourteen different SAv-metal channels from one representative CHB donor. The same vial of PBMC from each donor was stained in parallel with the same 562-plex pMHC tetramers but two

different SAV-metal coding configurations as shown. (B) Magnitude of selected HBV-specific CD8+ T cells detected by highly multiplexed combinatorial pMHC tetramer strategy across various clinical stages of HBV infection. Plots show the frequency of antigen-specific CD8+ T cells for fifteen predictive HBV epitope clusters and six representative known control viral epitopes (shaded box). Epitope sequences in bold face indicate previously unpublished sequences. \* means this epitope cluster contains more than one peptide (related to Supplementary Table 1). Dash lines on the y-axis are 0.002.

Fig. 9 The overall magnitudes of antigen-specific CD8+ T cells response in various clinical stages during HBV infection.

(A) Upper panel, the total number of different epitopes derived from four hepatitis B viral proteins (envelope, polymerase, core and x) detected in each individual patient across various clinical stages. Lower panel, the sum of frequency (%) of every detected antigen-specific CD8+ T cells for four different hepatitis B viral proteins in each individual patient across various clinical stages. n.s. = no significance.

Fig. 10 The validation of highly multiplexed combinatorial pMHC tetramer strategy in HLA-A\*11:01 and non-HLA-A\*11:01 donors.

(A) Cells from each donor were stained with selected 120-plex (Supplementary Table 1) pMHC tetramers coded with three different SAV-metal using two different sets of coding configurations. Experiments were performed independently and cells were stained and gated on live CD3+, Dump-(CD4+CD19+CD16+) and CD8+. Bar graphs indicate the frequencies for each epitope. (B) Representative dot plots show the pMHC tetramer positive cells and their signals of coded SAV-metals by different combinations of nine metal-tag SAV.

Fig. 11 The validation and reproducibility of antigen-specific CD8+ T cells using flow cytometry and the serological measurement of healthy donors.

(A) Correlations of the detected frequencies between FACS (single fluorochrome-coded) versus CyTOF (combinatorial metal-coded) based experiments. Each dot is one individual patient. (B) Representative FACS dotplots of selected HBV epitopes. The numbers are the frequencies of total CD8+ T cells. (C) Left, the HBsAb (anti-HBsAg antibody) titers in healthy donors (HD). Right, the frequency of HBVpol387-specific CD8+ T cells in HD in different HBV

serological (HBsAb, HBcAb and HBeAb) status. The levels of serum antibodies against different HBV viral antigens were measured by ELISA. Red circle indicates the only individual who was tested positive for HBcAb and HBeAb.

5 Fig. 12 In vitro expansion of antigen-specific CD8<sup>+</sup> T cells upon peptide stimulation.

(A) PBMC from different patient groups were expanded by corresponded viral peptides for 10 days. The frequency of antigen-specific CD8<sup>+</sup> T cells were determined as the same as ex vivo pMHC tetramer staining experiment. The number above each graph indicates the significant p value.

10

Fig. 13 The expression levels of nine different inhibitory receptors on antigen-specific CD8<sup>+</sup> T cells.

(A) The expression level of inhibitory receptors on selected antigen-specific CD8<sup>+</sup> T cells. The numbers on x-axis indicate the peptide cluster number for each specificity (epitopes) (see table\_S1). Colored dots are the four selective HBV epitopes (090\_HBV-P-282, 106\_HBV-P-387, 178\_HBV-C-169 and 283\_HBV-C-195v2). Color legends indicate the different HBV clinical stages. Shaded areas are control viral epitopes. Statistical significances were only shown for selected epitopes. p value less than 0.0001 (short ticks) or other values (long ticks) are indicated.

20

Fig. 14 The epitope frequencies in a longitudinal patient cohort of HBeAg-seroconverters.

(A) Deep sequencing analysis of HBV viral DNA showed the different dynamic of viral mutations on selective epitopes identified by highly multiplexed combinatorial pMHC tetramer strategy. A longitudinal patient cohort included treatment naive CHB patients who spontaneously underwent HBeAg-seroconversion (S, bottom) or Non-seroconversion (C, upper) across the similar time frame in chronological order. HBVpol387 (LVVDFSQFSR) proportion (upper right) in the viral population sequenced at each time point, out of a maximum of 1. This epitope remained fixed in all patients and no change was observed. Epitope ID and sequence was as listed in table S1. Detailed epitope mutation data can be found in table S4.

30

Fig. 15 Cellular profiles of subset clusters of HBV-specific CD8<sup>+</sup> T cells identified by Phenograph and the enrichment strategy.

(A) Subset enrichment strategy for eight major Phenograph subset clusters of HBV-specific CD8<sup>+</sup> T cells. Gating strategy was defined in a representative experiment and defining markers for each subset clusters were then applied onto three different batches of experiments to identify the proportion of each subset cluster within HBVpol387 and HBVcore169-specific CD8<sup>+</sup> T cells. Shaded areas are the target clusters color-coded as in Figure 2A. Black dots are cells in the target cluster. Grey dots are other antigen-specific CD8<sup>+</sup> T cells. The numbers indicate the frequencies.

Fig. 16 Unsupervised Phenograph clustering analysis identified multifactorial T cell heterogeneity of HBV-specific CD8<sup>+</sup> T cells.

(A) Representative plots of cellular cluster showed HBVpol387(LVVDFSQFSR)-specific CD8<sup>+</sup> T cells were enriched in distinct regions across patient groups. Related to Fig. 2A-B. Black dots are HBVpol387-specific CD8<sup>+</sup> T cells from indicated donor. Grey dots are the combined of all antigen-specific CD8<sup>+</sup> T cells of twenty individuals including all patient groups. The numbers indicate the proportion of highlighted clusters within HBVpol387-specific CD8<sup>+</sup> T cells. (B) The proportion of three cellular clusters (C7, C11 and C14) in HBVpol387-specific CD8<sup>+</sup> T cells. The proportion of cellular cluster C2 in HBVcore169-specific CD8<sup>+</sup> T cells. (C) Unsupervised Phenograph clustering showed the phenotypic difference of HBVpol282 and HBVcore195. Stacked bar charts showed the distribution of nineteen cellular clusters (Fig. 2A) in individual across patient groups. (D) pMHC tetramer intensity was quantified by averaging the median numbers of metals of four different SAV-metals coded on tetramer positive cells. Plots show the normalized value (z-score) of tetramer intensity on different selected virus-specific CD8<sup>+</sup> T cells (left), and HBVcore169-specific CD8<sup>+</sup> T cells (right) across various patient groups. (E) The proportion of cellular cluster and the expressions of cellular markers of HBVcore169-specific CD8<sup>+</sup> T cells derived from one IT patient (IT07), whose frequency (0.00193%) was just below the imposed cut-off. (F) Hierarchical clustering of cellular markers expression of EBVEBNA3B-specific CD8<sup>+</sup> T cells from the same individuals as Fig. 1D across patient groups.

Fig. 17 The co-expressions of inhibitory receptors on virus-specific CD8<sup>+</sup> T cells.

(A) The average numbers of co-expressed inhibitory receptors on different antigen-specific CD8+ T cells in individuals. Plots were combined of four independent experiments without the measurement of cellular marker TIGIT. Each dot is one individual. (B) The average numbers of co-expressed inhibitory receptors on different antigen-specific CD8+ T cells in individuals. Plots were from an experiment with the measurement of all nine inhibitory receptors.

Fig. 18 The heterogeneous multi-functional subsets of virus-specific CD8+ T cells.

(A) The detailed One-SENSE functional clusters of different antigen-specific CD8+ T cells across patient groups. n=4~5 per patient group. Functional subsets were labelled as in Fig. 4A. (B) The correlation of co-producing Granzyme A and K with the co-expression of 2B4 and TIGIT on virus-specific CD8+ T cells. The proportion of Non-functional subset (black) was correlated to the sustained expression of HVEM on virus-specific CD8+ T cells. Dots were the different virus-specific CD8+ T cells from individual patients. (C) The representative contour plots of HBVenv304-specific CD8+ T cells showed the heterogeneous multi-functionality between patients.

Fig. 19 Diverse characteristics of epitope-specific TCR $\beta$  repertoire using TCRdist.

(A) Nine TCR motif clusters identified by Phenograph were presented, and the representative TCR motifs were shown using average-linkage dendrogram using TCRdist algorithm. Related to Fig. 5. (B) TCRdiv diversity measures of total CD8+ T cells TCR $\beta$  repertoire between patient groups are shown. (C) The length (aa, amino acid) of CDR3 $\beta$  of total and epitope-specific CD8+ T cells across patient groups. Error bars are mean and SEM. Statistical analysis was calculated using Gaussian fit with the null hypothesis "one curve fits all groups".

Fig. 20 The dynamic of cellular response and viral mutation of HBVcore169-specific CD8+ T cells in a longitudinal patient cohort.

(A) The liver inflammation scores of two patients (HBeAg+03 and HBeAg-01) from a longitudinal patient cohort were shown. ALT, alanine aminotransferase. AST, aspartate aminotransferase. AFP, alpha-fetoprotein. ALP, alkaline phosphatase. Related to Fig. 6A. (B) PBMCs from pre- and post-treatment of HBeAg+03 and HBeAg-01 were used to resolve the

specific tetramer response for 7 different peptides in cluster 178 (see table S1). Cells were evenly split and stained with the corresponded (as indicated above) pMHC tetramer independently using flow cytometry. The numbers indicate the frequency of CD8+ T cells. Virus from paired serum samples were sequenced to determine the variants (in frequency) of the epitope (left). WT, wild type. The same sequences were color coded as indicated.

Fig. 21 The staining quality of cellular markers including nine inhibitory receptors using mass cytometry.

(A) Dotplots show the expression levels of markers probed on antigen-specific (left) versus global (right) CD8+ T cells in patient's PBMC. Upper panels are ex vivo staining. Lower panels are cells from in vitro peptide stimulation. For the better visualization of dotplots, all detected antigen-specific CD8+ T cells were pooled from 18 patients (left panel, n=4~5 per patient group, including IT, IA, InA and R). PD-1-expressing HBVcore169-specific CD8+ T cells from a CHB patient is also showed. Plots were from two independent experiments (ex vivo and in vitro).

The present invention will be described with respect to particular embodiments and with reference to certain drawings, but the invention is not limited thereto but only by the claims.

Any reference signs in the claims shall not be construed as limiting the scope. The drawings described are only schematic and are non-limiting. In the drawings, the size of some of the elements may be exaggerated and not drawn on scale for illustrative purposes.

As used herein, except where the context requires otherwise, the term "comprise" and variations of the term, such as "comprising", "comprises" and "comprised", are not intended to exclude further additives, components, integers or steps. As used herein, except where the context requires otherwise, "comprise" and "include", or variations of the term such as "including" can be used interchangeably.

Where an indefinite or definite article is used when referring to a singular noun e.g. "a" or "an", "the", this includes a plural of that noun unless something else is specifically stated. Furthermore, the terms first, second, third and the like in the description and in the claims,

are used for distinguishing between similar elements and not necessarily for describing a sequential or chronological order. It is to be understood that the terms so used are interchangeable under appropriate circumstances and that the embodiments of the invention described herein are capable of operation in other sequences than described or illustrated herein.

The following terms or definitions are provided solely to aid in the understanding of the invention. Unless specifically defined herein, all terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Press, Plainsview, New York (1989); and Ausubel et al., Current Protocols in Molecular Biology (Supplement 47), John Wiley & Sons, New York (1999), for definitions and terms of the art.

The definitions provided herein should not be construed to have a scope less than understood by a person of ordinary skill in the art.

## EXAMPLE

### Materials and Methods

#### 1. Patient samples and PBMC isolation

Patients (table S3) with HBV infection were recruited with fully informed written consent from Division of Gastroenterology and Hepatology at National University Health System, Singapore. The respective local ethical institutional review boards approved the study, and the recruitment and sampling of suitable patients was completed at hospital. Up to 60 ml of blood was taken, peripheral blood mononucleated cells (PBMC) were further isolated by using Ficoll-separation (Ficoll-Paque PLUS, GE Healthcare). All patients had clinical, serological and virological evidences of chronic hepatitis B infection with detection of HBsAg and HBV DNA, and no positive result for the presence of HIV-1 and -2, and HCV. Three CHB patient groups, Immune Tolerant (IT, HBV DNA>2000 IU/ml, ALT<40 IU/ml, HBeAg+),

Immune Active (IA, HBV DNA>2000 IU/ml, ALT>40 IU/ml, HBeAg+), Inactive Carrier (InA, HBV DNA<200 IU/ml or undetectable, ALT<40 IU/ml, HBeAg-) and one group of acute resolved patients (R, undetectable HBV DNA, HBsAg- and anti-HBc antibody+) were enrolled in this study. Each CHB patient had at least three adjacent time points indicating consistent virological and serological evidences for the referring clinical stages. All patients were treatment free from any antiviral drug or clinical intervention at the time of blood draw. Patients received Entecavir (ETV) were followed up longitudinally and enrolled. Serological and virological scores (serum HBV DNA, HBeAg and HBsAg) and liver function test were determined by clinical laboratory at hospital or ELISA. Blood from anonymous healthy donors were recruited under Singapore Immunology Network (SIgN) institutional review board. Healthy cord blood samples were purchased from Singapore Cord Blood Bank under the institutional review board, without detection of HIV-1 and -2, HTLV-I and II, HCV, CMV, HBsAg and anti-HBc antibody. HLA-A\*11:01 was confirmed by typing service from BGI Genomics.

Supplementary Table 3: List of the patients samples and the clinical and serological information.

| Group types                | Patient group    | n  | %    | Age          | 100% CD4 <sup>+</sup> (ng/ml) | 40% CD4 <sup>+</sup> | 40% CD4 <sup>+</sup> | 100% CD4 <sup>+</sup>     | 100% CD4 <sup>+</sup>     | 100% CD4 <sup>+</sup> | 100% CD4 <sup>+</sup> |
|----------------------------|------------------|----|------|--------------|-------------------------------|----------------------|----------------------|---------------------------|---------------------------|-----------------------|-----------------------|
| PMHC                       | Non-HIV (n=10)   | 11 | 0%   | 72.5 (26-76) | 4.4e+07 (1.2e+07-8.5e+07)     | 50 (21-100)          | 25 (5-100)           | 5.1e+07 (5.3e+07-8.2e+07) | 5.6e+07 (5.3e+07-1.1e+08) | 948 (669-1000)        | 4.1 (0-12.0)          |
|                            | Non-HIV (n=10)   | 20 | 0%   | 63.5 (39-68) | 6.6e+07 (2.3e+07-1.1e+08)     | 81 (40-100)          | 80 (25-100)          | 1.3e+07 (5.3e+07-4.8e+07) | 1.6e+07 (3.3e+07-1.2e+08) | Not detected          | 4.5 (0-19.8)          |
|                            | Non-HIV (n=10)   | 15 | 0%   | 62.5 (34-72) | 6.6e+07 (2.3e+07)             | 29 (0-100)           | 23 (5-100)           | 4.8e+07 (0-1.6e+08)       | 1.6e+07 (0-1.6e+08)       | 0 (0-100)             | 140.1 (16.7-142.5)    |
|                            | Acute HIV (n=10) | 13 | 0%   | NA           | Not detected                  | 0%                   | 0%                   | 0%                        | 0%                        | 0 (0-100)             | Not detected          |
|                            | Steady State HIV | 12 | 0%   | 34 (25-42)   | NA                            | 0%                   | 0%                   | 0%                        | 0%                        | 0 (0-100)             | 0 (0-100)             |
| HIV (n=10)                 |                  | 10 | 0%   | 63%          | 0%                            | NA                   | NA                   | NA                        | NA                        | NA                    | NA                    |
| Longitudinal patient group |                  | n  | %    | n of samples | 100% CD4 <sup>+</sup> (ng/ml) | 40% CD4 <sup>+</sup> | 40% CD4 <sup>+</sup> | 100% CD4 <sup>+</sup>     | 100% CD4 <sup>+</sup>     | 100% CD4 <sup>+</sup> | 100% CD4 <sup>+</sup> |
| PMHC                       | 100%             | 8  | 100% | 5-7          | 6.6e+07 (2.3e+07)             | 73 (25-100)          | 28 (5-100)           | Positive                  | Negative**                | NA                    | Negative**            |
|                            | 40%              | 6  | 87%  | 4-8          | 6.6e+07 (2.3e+07)             | 18 (5-100)           | 24 (5-100)           | Positive                  | Positive                  | NA                    | Negative              |
| Longitudinal patient group |                  | n  | %    | n of samples | 100% CD4 <sup>+</sup> (ng/ml) | 40% CD4 <sup>+</sup> | 40% CD4 <sup>+</sup> | 100% CD4 <sup>+</sup>     | 100% CD4 <sup>+</sup>     | 100% CD4 <sup>+</sup> | 100% CD4 <sup>+</sup> |
| Serum                      | 100% (n=10)      | 8  | 0%   | 4-8          | 8.1e+07 (6.6e+07)             | 50 (21-100)          | 50 (21-100)          | 50 (21-100)               | 50 (21-100)               | 50 (21-100)           | 50 (21-100)           |
|                            | 40% (n=10)       | 7  | 0%   | 5-7          | 8.1e+07 (6.6e+07)             | 50 (21-100)          | 50 (21-100)          | 50 (21-100)               | 50 (21-100)               | 50 (21-100)           | 50 (21-100)           |

Table S3

5

## 2. 562-plex combinatorial (Quadruple/Triple SAV-metal coded) pMHC tetramers

Fourteen different SAV-metals were made by labelling streptavidin with fourteen different metal isotopes. Similar to previous work reported in E. W. Newell *et al.*, Combinatorial tetramer staining and mass cytometry analysis facilitate T-cell epitope mapping and characterization, each SAV-metal was diluted into 20 µg/ml in EDTA-free W-buffer on the same day of tetramerization of pMHC. Two different configurations of quadruple SAV-metal coding for 562-plex pMHC tetramers were generated using a R-based script for a 14-choose-4 scheme (for 1001 combinations). The script was then loaded onto TECAN Freedom EVO200 automatic liquid distribution robot to prepare the designed combinations of quadruple SAV-metal mixtures (each mixture contains 4 different SAV-metals) in 2ml 96-well deep well plates. To form pMHC tetramers, each peptide-loaded HLA-A\*1101 monomer (562 different pMHC monomers) was randomly assigned for four different SAV-metals. To reach a 1:4 ratio of streptavidin to pMHC, quadruple SAV-metal mixtures were added to the corresponded pMHC monomer in a stepwise manner of four additions, each has 10 min incubation at room temperature. 10 µM D-biotin was added into the reaction at the end for another 10 min at room temperature to saturate unbound streptavidins. The 562-plex pMHC tetramers were combined and concentrated down to 5 µg/ml per pMHC tetramer in 10% FBS CyFACS buffer using Amicon 50kDa cut-off concentrator (Millipore). The total amount of protein in each concentrator was limited to 300 µg. To reach the desired

25

concentration and volume, multiple spins were performed at 700 xg for 5 min. The 562-plex tetramers were then filtered by 0.1µm filter tube (Millipore) at 2000 xg for 25 min. The cocktail of tetramers was kept on ice, and then spin at 14,000 xg for 1 min in a 1.5 ml Eppendorf tube to remove the remaining aggregates prior the staining.

5

The combinatorial streptavidin codings were re-scrambled for every independent 562-plex combinatorial pMHC tetramer staining experiment.

10

For selected experiments, a 9-choose-3 (84 combinations) or 8-choose-3 (56 combinations) scheme were used to cover 120-plex (40 peptide clusters) or 50-plex (17 peptide clusters) combinatorial triple coded pMHC tetramers staining preferentially selected (table S1) for more phenotypic analysis, or in vitro peptide stimulation (table S2).

Supplementary Table 1. List of domain 4B-A-AC111G-identified epitopes and the frequency across various patient groups

| Antigen Name      | Epitope      | Sequence   | Frequency in 1st group | Frequency in 2nd group |     |     |     | Frequency in 3rd group |     |     |     |
|-------------------|--------------|------------|------------------------|------------------------|-----|-----|-----|------------------------|-----|-----|-----|
|                   |              |            |                        | 1st                    | 2nd | 3rd | 4th | 1st                    | 2nd | 3rd | 4th |
| Protein B (cont.) | 1-1000-1-1   | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 2-1000-1-1   | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 3-1000-1-1   | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 4-1000-1-1   | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 5-1000-1-1   | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 6-1000-1-1   | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 7-1000-1-1   | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 8-1000-1-1   | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 9-1000-1-1   | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 10-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
| Protein B (cont.) | 11-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 12-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 13-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 14-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 15-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 16-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 17-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 18-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 19-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 20-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
| Protein B (cont.) | 21-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 22-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 23-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 24-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 25-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 26-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 27-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 28-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 29-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 30-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
| Protein B (cont.) | 31-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 32-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 33-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 34-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 35-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 36-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 37-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 38-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 39-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 40-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
| Protein B (cont.) | 41-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 42-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 43-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 44-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 45-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 46-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 47-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 48-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 49-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 50-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
| Protein B (cont.) | 51-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 52-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 53-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 54-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 55-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 56-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 57-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 58-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 59-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 60-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
| Protein B (cont.) | 61-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 62-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 63-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 64-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 65-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 66-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 67-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 68-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 69-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 70-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
| Protein B (cont.) | 71-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 72-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 73-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 74-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 75-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 76-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 77-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 78-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 79-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 80-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
| Protein B (cont.) | 81-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 82-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 83-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 84-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 85-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 86-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 87-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 88-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 89-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 90-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
| Protein B (cont.) | 91-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 92-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 93-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 94-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 95-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 96-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 97-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 98-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 99-1000-1-1  | 1000000000 |                        |                        |     |     |     |                        |     |     |     |
|                   | 100-1000-1-1 | 1000000000 |                        |                        |     |     |     |                        |     |     |     |

Table S1

Table S1 - continued

Table S1 - continued

| PBMC, ex vivo panel 1, 145Axi-choose-4 |                   |        |                  |      | PBMC, ex vivo panel 2, 90Axi-choose-3 |                   |                  |         |      | PBMC, ex vivo panel 3, 80Axi-choose-3 |                   |                  |         |      |
|----------------------------------------|-------------------|--------|------------------|------|---------------------------------------|-------------------|------------------|---------|------|---------------------------------------|-------------------|------------------|---------|------|
| Antibody                               | Antibody on label | Clone  | Company          | ISOT | Antibody                              | Antibody on label | Clone            | Company | ISOT | Antibody                              | Antibody on label | Clone            | Company | ISOT |
| Y-89                                   | CD14              | 1B10   | PharMingen (DCC) |      | CD14                                  | 1B10              | PharMingen (DCC) |         |      | CD14                                  | 1B10              | PharMingen (DCC) |         |      |
| Rb-102                                 | Cell barcode      |        |                  |      | Cell barcode                          |                   |                  |         |      | Cell barcode                          |                   |                  |         |      |
| Rb-103                                 | DNA barcode       |        | PharMingen (DCC) |      | Live/Dead                             |                   |                  |         |      | Cell barcode                          |                   |                  |         |      |
| PS-104                                 | Cell barcode      |        |                  |      | Cell barcode                          |                   |                  |         |      | Cell barcode                          |                   |                  |         |      |
| PS-105                                 | Cell barcode      |        |                  |      | Cell barcode                          |                   |                  |         |      | Cell barcode                          |                   |                  |         |      |
| PS-106                                 | Cell barcode      |        |                  |      | Cell barcode                          |                   |                  |         |      | Cell barcode                          |                   |                  |         |      |
| PS-108                                 | Cell barcode      |        |                  |      | Cell barcode                          |                   |                  |         |      | Cell barcode                          |                   |                  |         |      |
| PS-110                                 | Cell barcode      |        |                  |      | Cell barcode                          |                   |                  |         |      | Cell barcode                          |                   |                  |         |      |
| CD-112/114                             | Q-BB95-CD114      | TsB4   | Molecular Probes |      | CD100/CD114                           | TsB4              | Molecular Probes |         |      | CD100/CD114                           | TsB4              | Molecular Probes |         |      |
| Is-113                                 | Empty             |        |                  |      | Empty                                 |                   |                  |         |      | Empty                                 |                   |                  |         |      |
| Is-115                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-119                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-140                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-141                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-142                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-143                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-144                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-145                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-146                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-147                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-148                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-149                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-150                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-151                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-152                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-153                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-154                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-155                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-156                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-157                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |
| Is-158                                 | CD137             | 1C3237 | Biogen           |      | CD137                                 | 1C3237            | Biogen           |         |      | CD137                                 | 1C3237            | Biogen           |         |      |

5

### 3. Highly multiplexed pMHC tetramer, antibody staining and CD8 T cell enrichment

Cryopreserved PBMC were thawed and washed with complete RPMI (10% FBS + 1% penicillin/streptomycin/L-glutamine + 1% 1M HEPES) (Gibco, Invitrogen), and rest for 3 hours at 37°C. After the recovery, cells were harvested and seeded on a non-treated 96-well plate, and about 10 million cells per patient were used and split evenly in two separated wells for two configurations of 562-plex combinatorial pMHC tetramer staining. 50 µM dasatinib was incubated with cells for 30 min at 37°C, 5% CO<sub>2</sub>, to prevent the downregulation of TCR (30). Cells were washed with CyFACS buffer (2 mM EDTA + 0.05% sodium azide + 4% FBS in PBS) and incubated with 200 mM cisplatin (Pt-195) for 5 min on ice, or rhodium (Rh-103) for 20 min at room temperature (table S2) for viability measurement. After wash once with CyFACS buffer, cells from the same donor in separated wells were stained with 50 µl of cocktail containing the same 562-plex pMHC tetramers but completely different SAv-metal coding configurations for 1 hour in room temperature in the presence of 1:100 Fc block (Biolegend). Cells were washed twice with CyFACS buffer after incubation, and resuspend in 50 µl of T cells or CD8 T cells enrichment kit (STEMCELL) antibody cocktail in 1:10 in CyFACS buffer for 30 min on ice. Cells were then washed, and stained with 50 µl of primary antibody cocktail (table S2 and Fig. 21) for 30 min on ice. Excessive antibodies were removed by washing the cells twice with CyFACS buffer, and cells were resuspended with 4 µl of enrichment beads (STEMCELL) + 46 µl of CyFACS buffer for 15 min on ice. After the staining, cells were washed with PBS and fixed with 200 µl of 2% PFA (paraformaldehyde, Electron Microscopy Sciences) overnight at 4°C. On the next day, PFA was removed and cells were incubated with permeabilization buffer (Biolegend) at room temperature for 10 min, and then resuspended with 50 µl of intracellular antibody cocktail for 30 min at room temperature. For subsequent dual mass-tag cellular barcoding, 2 mM bromoacetamidobenzyl-EDTA (BABE; Dojindo) with 0.5 mM PbCl<sub>2</sub> was dissolved in HEPES buffer, and each sample was given a unique combination of metal-barcode (BABE-Pd-102, BABE-Pd-104, BABE-Pd-106, BABE-Pd-108, BABE-Pd-110) on ice for 30 min. After 5 min incubation with CyFACS buffer on ice, cells were labeled by Iridium DNA interchelator (Ir-191/193, Fluidigm DVS) in 2% PFA at room temperature for 20 min. Cells were then washed with CyFACS buffer and CD8 T cells were negatively selected using EasySep™ Magnet

(STEMCELL) according to manufacturer's instruction. Enriched cells were washed twice by MilliQ water and ready for mass cytometry acquisition.

#### 4. Statistical analysis

Non-parametric analysis of variance (ANOVA) was used for group comparison unless indicated elsewhere.  $p < 0.05$  by non-parametric ANOVA allowed the subsequent multiple comparison test. P values were calculated using Prism software (GraphPad). All error bars are median and SEM.

#### 5. Amplification of HBV genome and library construction

Seven treatment-naïve HBeAg non-seroconverters and eight HBeAg seroconverters of patients chronically infected by HBV (including genotype B and C) were recruited in National University Health System, Singapore. Multiple longitudinal serum samples (5 to 15 time points per patient) from each patient were taken across the event of HBeAg seroconversion. Deep sequencing analysis was performed in all serums samples by sequencing the whole HBV viral genome. Similar to previously description, primers (5'-GCTCTTCTTTTTCACCTCTGCCTAATCA-3' and 5'-GCTCTTCAAAAAGTTGCATGGTGCTGG-3') were used to generate full-length amplicons of the HBV genome. Polymerase chain reaction (PCR) was performed using the PfuUltra™ II Fusion HS DNA Polymerase (Stratagene, La Jolla, California, USA) according to the manufacturer's instructions. The 3.1 kb fragment was extracted from 1% agarose gel prepared in 1 x TBE buffer, using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA) and the concentration of the extracted product was measured using the NanoDrop ND 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Each sample was fragmented into 100-300 bp using the Covaris S2 (Covaris, Woburn, MA, USA). (Shearing conditions - Duty cycle: 20%; Intensity: 5; Cycles per burst: 200; Time: 110 seconds). After fragmentation, the samples were purified using the QIAquick PCR purification kit (Qiagen, Valencia, CA, USA). The DNA 1000 Chip was used with the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) to check the size and quality of the fragmented products. For library construction, the KAPA Library Preparation Kit (KAPA Biosystems) was used according to the manufacturer's instructions. The library

construction includes end repair, A-tailing, ligation of adapters and a final PCR step that incorporates the indexes into the samples. Illumina TrueSeq adapters and indexes were used (Illumina, San Diego, CA, USA). PfuUltra™ II Fusion HS DNA Polymerase was used for this final PCR step according to the manufacturer's instructions. The samples were then  
5 cleaned up using the Agentcourt AMPure XP (Beckman Coulter) on a 1:1 ratio of beads to sample. To check on the size and concentration of the ligated products, the 2100 Bioanalyzer with DNA 1000 Chip was used. The quality and quantity of the products were determined by running a quantitative PCR. The reactions were prepared using the KAPA Library Quantification kit (KAPA Biosystems) and the run was done on the LightCycler 480 II  
10 real time thermal cycler (Roche Applied Science, Indianapolis, IN, USA) according to the manufacturer's instructions. Samples were sequenced in the Genome Institute of Singapore on the Illumina HiSeq 2500 to obtain multiplexed 101 bp paired-end reads.

#### 6. SNV analysis on viral epitope

We modified the reference genome P121214 such that ambiguous positions (e.g. R/W/Y) were replaced by one of the relevant bases (A/C/G/T) randomly. 101bp paired -end reads were mapped to the modified reference genome with BWA-MEM version 0.7.10-r789 and Single-nucleotide variant (SNV) calling was carried out with LoFreq version 2.1.2. Coverage  
20 depth averaged around 104~105 for all samples. SNVs were filtered by frequency (>5%), SNV quality (>1000), and coverage depth (>100) to remove false positives. SNVs within PCR primer regions were also ignored due to high error rates. SNVs passing quality filters were then sorted and only non-synonymous mutations that changed >20% in frequency between the early and late time points were kept (custom per script). These SNVs were candidates  
25 for adaptive epitopes, and were matched to known epitope sequences for further tetramer experiments.

#### 7. HBV epitope prediction and peptide synthesis

The consensus sequences of HBV from each patient were further determined and translated into amino acid sequence for each open reading frame (ORF). The predicted binders (peptides) restricted to HLA-A\*1101 were then generated by NetMHC software (v3.4 server,  
30

<http://www.cbs.dtu.dk/services/NetMHC/>) based on the consensus sequences derived from HBV proteins (core, polymerase, x and envelope), including all possible binding variants for 8-, 9-, 10-, and 11-mer peptides that above the binding threshold (for score > 0.4 predicting weak binding and score > 0.6 predicting strong binding). The prediction scheme produced 484 unique HLA-A\*1101-restricted HBV epitopes, together with 78 known HLA-A\*1101-restricted epitopes derived from other pathogens or self-proteins, a total of 562 different pMHC tetramers (table S1) were therefore made from these peptides for subsequent highly multiplexed combinatorial pMHC tetramer mapping. Immune Epitope Database (IEDB) was used to report previously unidentified epitope sequences. All peptides were synthesized by Mimotopes (Australia) with purity > 85%.

#### 8. Peptides sequence similarity and cluster assignment

To avoid the incorrect interpretation from cross-reactive T cell epitopes in the 562-plex pMHC library that comprised of all viral proteins and variants, sequences of the library were pooled and loaded onto a Biostrings-based R-written environment. Similar to BLAST (Basic Local Alignment Search Tool), the biological sequence and matching algorithm performed pairwise alignment to calculate the peptide binding score based on their sequence similarity. Total of 284 peptide clusters were assigned, and the peptides in each cluster are listed (table S1). Peptides within the same cluster were then given the same quadruple SAV-metal coding for highly multiplexed combinatorial pMHC tetramer strategy.

#### 9. Generation of HLA-A\*11:01 monomer and 562-plex pMHC library

The inclusion bodies of HLA-A\*1101 were produced (76, 77), and refolded with a UV-cleavable peptide H-RVFA(J)SFIK-OH, where J is ANP (3-Amino-3-(2-nitrophenyl) propionic acid) linker. The protein was purified and biotinylated, and stored in PBS + 50% glycerol at -20°C. Peptide exchange was performed at 0.1 mg/ml of HLA-A\*1101 monomer in 100 µl PBS with 25 µM of peptide of interest in a 96-well plate. The reaction was exposed to 365 nm UV irradiation for 5 min twice using UVP CL-1000 Ultraviolet Crosslinker, the plate was further sealed and stored at 4°C overnight to complete the exchange.

## 10. Streptavidin (SAv) production and metal labeling

Streptavidin with free cysteines residues separated by glycine linkers were used for recombinant expression. Briefly, purified streptavidin was made in-house and stored in 10 mM TCEP in 20 mM HEPES (pH 7.2) buffered saline as frozen aliquots at -80°C. After conjugation using DN3 polymer labeling kits and filtering using 0.1 µm filters (Amicon), the metal-tag streptavidin conjugates (SAv-metal) were transferred to a new 30kD concentrator (Merck) to perform five washes with EDTA-free W-buffer. SAv-metal was adjusted to final concentration at 200 µg/ml prior the formation of tetrameric pMHC complex.

## 11. Antibody-metal conjugation

Purified antibodies without carrier proteins were purchased as listed (table S2). 50 or 100 µg of antibody was conjugated with metal-attached maleimide-coupled DN3 MAXPAR (Fluidigm DVS) chelating polymer according to manufacturer's instruction (Fluidigm DVS) as previously described. All metal isotopes were purchased from Fluidigm DVS or TRACE Sciences International Inc. as listed (table S2).

## 12. Mass cytometry and data pre-processing

All experiments were acquired by CyTOF2 (Fluidigm DVS) systems. Cells were wash by MilliQ water twice, filtered, and immediately acquired by mass cytometry with an acquisition rate of 300~350 cells/sec. 2% of Four EQ beads (Fluidigm DVS) were mixed with cell suspension. To normalize signal variations of CyTOF2, the output FCS files were normalized based on the added beads that has been previously described. Normalized FCS files were further loaded onto a Unix-based R-written script and all zero values were randomized into values between 0 to -1 using uniform distribution.

## 13. Self-validated automatic deconvolution of antigen-specific T cells

After data pre-processing, lived CD8+ T cells were gated using FlowJo v9.7.6 (Tree Star Inc.) and individual samples were de-barcoded based on the dual mass-tag cellular barcodes

using Boolean gates. Two SAV-metal coding configurations from the same donor were barcoded and exported independently. For optimal automatic identification of tetramer positive cells, multiple safety parameters and thresholds were built and subjectively defined using an R-written script, respectively (Fig. 1 and Fig. 7). Briefly, thresholds for each SAV-metal channel were manually defined by gating a tetramer negative population for all fourteen SAV-metal channels. Based on thresholds (Threshold  $X=T_x$  and Threshold  $Y=T_y$ ) of every SAV-metal channel, the safety factors then objectively identify the tetramer positive population using the pre-set geometric criteria ( $Y/X \text{ Slope}=k$ ,  $X/Y \text{ Slope}=k$ , and  $\text{Width}=w$ ) (Fig. 7). All antigen-specific CD8<sup>+</sup> T cells identified by highly multiplexed combinatorial pMHC tetramer strategy in this report have to firstly pass both the thresholds and auto-gating stringency parameters. Secondly, the corresponded four SAV-metals coded on each pMHC tetramer must have exclusive and higher metal intensity than rest of the ten SAV-metal channels. The deconvolution algorithm excluded any tetramer positive cells that have less, or more than four SAV-metals coding. The tetramer positive cells identified in two different SAV-metal coding configurations of the same donor were further calculated for their signals correspondence by using statistical simulation (Fig. 7), with  $p<0.05$  was considerate a confident detection. Finally, antigen-specific CD8<sup>+</sup> T cells who passed all the above-mentioned tests with frequency  $>0.002$  of total CD8<sup>+</sup> T cells were selected for further high-dimensional data analysis.

#### 14. High-dimensional cytometric data visualization

Validated antigen-specific CD8<sup>+</sup> T cells were exported individually from each donor for dimensionality reduction analysis. Detailed methodology for t-SNE and One-SENSE can be found elsewhere. Briefly, t-SNE and One-SENSE were performed using custom R scripts based on “flowCore” and “Rtsne” packages downloaded from The Comprehensive R Archive Network (CRAN). All data were transformed using the “logicleTransform” function and  $w=0.25$ ,  $t=16409$ ,  $m=4.5$ ,  $a=0$  as input parameters to roughly match scaling historically used in FlowJo. Cellular markers analyzed by t-SNE and One-SENSE were indicated (table S2). For One-SENSE, cellular markers in each T cell category (“Differentiation+TNFR”, “Exhaustion” and “Trafficking”) were subjectively assigned for categorical analysis. Aligned heatplots represent the distribution of marker positive cells in percentage on each “bin” on the axis

(category) constructed by cells residing in small ranges of values. Positive population of markers was manually defined and markers of the same category were combined for each dimension using 250 bins.

- 5 The 3D visualization of One-SENSE was built from numerous consecutive 3D images supported by “rgl” package based on the three dimensions of One-SENSE analysis. The continuous image sequences were subsequently combined by Sequimago (AppleScript) to generate 3D movies.
- 10 Logistic regression was performed using the drc R package (v 3.0.1) using a 3-parameters logistic model. Support vector machines was trained on the 7 parameters common across the two datasets using the e1071 R package (v 1.6-8) using default parameters (epsilon-regression with radial kernel, gamma of 1/7 and epsilon of 0.1).

## 15 15. Flow cytometry and cell sorting

- Cells were prepared as the same fashion as mass cytometry experiments. After incubation with dasatinib, cells were washed with PBS and incubated with 50 µl of Live/Dead-Pacific Orange (Thermo Fisher) on ice for 20 min in dark. For single fluorochrome-tag pMHC
- 20 tetramers, peptide-exchange and tetramer formation were done using the same method as metal-tag pMHC tetramers in dark. PE-SAv (eBioscience), PE-Cy7-SAv (eBioscience), PE-Cy5-SAv (eBioscience), BV650-SAv (BD) and APC-SAv (Biolegend) were diluted to 20 µg/ml in PBS and added into pMHC monomer loaded with different peptides in the same manner as mentioned above. Cells then washed, and stained by tetramer cocktail in the same
  - 25 condition as mass cytometry experiment in dark. After two washed with FACS buffer, cells were stained with primary antibodies, Pacific Blue-CD14 (Biolegend), Pacific Blue-CD16 (Biolegend), Pacific Blue-CD19 (Biolegend), Alexa Fluor 700-CD3 (Biolegend), FITC-CD4 (Biolegend) and QD605-CD8 (Thermo Fisher) in FACS buffer for 30 min on ice in dark. Cells were then washed twice, filtered, and analyzed by LSRFortessa (BD). Fluorochrome-tag
  - 30 tetramer positive cells were stained using the same method and live-sorted using Aria II 5 lasers system (BD).

## 16. TCR high-throughput sequencing and spectratyping

Five populations, including four different tetramer-stained virus-specific CD8<sup>+</sup> T cells plus the total CD8<sup>+</sup> T cells from each donor were live-sorted, and genomic DNA were freshly extracted using Qiagen Blood & Tissue kit according to manufacturer's instruction. Cells from three donors per patient group, with five groups and 75 samples in total, were used for TCR sequencing. Briefly, TCR $\beta$  chains were amplified using a bias-controlled two-step multiplex PCR by ImmunoSEQ platform (Adaptive Biotechnologies). The first PCR amplified the CDR3 region of sorted T cells, followed by adding the adaptor sequences in second PCR, and subsequently sequenced by next-generation sequencing (NGS). Productive reads were generated from the reduction of amplification and sequencing bias. 1000, 5000, 1500, 2000 and 10000 cells were live sorted for HBVpol282-, HBVpol387-, HBVcore169-, HBVcore195-specific and total CD8<sup>+</sup> T cells, respectively.

TCR sequences were then exported from ImmunoSEQ Analyzer and the genes were defined by IMGT/HighV-Quest nomenclature. CDR3 $\beta$  length analysis was calculated by Prism using Gaussian fit with the null hypothesis "one curve fits all". The null hypothesis was rejected with statistical significance.

## 17. TCRdist measurements for quantifying epitope-specific repertoires

Quantifiable distances of epitope-specific TCR $\beta$  (TCRdist) were computed to be the similarity-weighted Hamming distances between the contacting regions of two TCR's CDR1, 2 and 3 as recently developed (53). TCRdist pipeline was downloaded (<https://github.com/phbradley/tcr-dist>) and installed in Python. TCR $\beta$  sequences were loaded and executed with additional command "--make\_fake\_alpha" and "--make\_fake\_qual". The unsupervised visualization of t-SNE maps were constructed based on the kernel PCA coordinates and the TCRdist distance matrix of TCR clones generated by TCRdist pipeline using a custom R-written script. The implementation for unsupervised clustering Phenograph algorithm, Rphenograph, was installed in R (<https://github.com/JinmiaoChenLab/Rphenograph>). A repertoire diversity metric (TCRdiv)

that generalizes Simpson's diversity index was used to measure the diversity of epitope-specific TCRs by accounting both the similarity and identity of the sequences.

#### 18. *In vitro* virus-specific CD8<sup>+</sup> T cells expansion

5

PBMC from patients were thawed and recovered, and then resuspend in AIM-V medium (with 2% human AB serum, 1% penicillin/streptomycin/L-glutamine, 1% 1M HEPES) (Gibco, Invitrogen) with 20 IU/ml of recombinant human IL-2 (R&D). Cells were pulsed with corresponded HBV or control peptides at 1  $\mu$ M per 1 million cells in 200  $\mu$ l medium in 96-well round-bottom tissue culture plate and cultured for 10 days at 37°C. Half of the medium was replaced as supplementary every three days without any peptide. On day 10, cells were restimulated with or without the corresponded peptides for 7 hours in the presence of Brefeldin A (eBioscience), monensin (eBioscience) and 0.5  $\mu$ g/ml anti-CD107a at 37°C. After incubation, cells were collected and then stained with the 50-plex pMHC tetramers (table S1) and surface antibodies. Intracellular cytokine staining was performed on the second day as indicated in (table S2). All staining, cellular barcoding, and CD8 T cell enrichment were done in the same manner as ex vivo staining described above.

10

15

#### 19. Enzyme-linked immunosorbent assay (ELISA)

20

Paired serum samples from patients were serially diluted in PBS, and level of HBeAg, HBsAg, HBeAb, HBsAb and HBcAb were determined using quantitative sandwich ELISA kits (Abnova and MyBioSource) according to manufacturer's instruction.

## 25 Results

### 1. Comprehensive HBV epitope mapping

30

To generate a comprehensive HBV targeting pMHC library, viral DNA was isolated and deep sequenced from serum samples of 15 longitudinal CHB patients to determine viral consensus sequences and common variants (Fig. 1A and Fig. 7A, see Materials and Methods). The sequences were loaded onto the NetMHC platform to predict possible A\*11:01-

restricted binders. 484 unique putative HBV epitopes above the predictive “weak binding” threshold were combined with 78 known epitopes derived from other common antigens to arrive at a total of 562 peptides, listed in table S1. Sequence homology was analyzed to group relatively similar peptides into the same cluster by a pairwise matching algorithm (Fig. 7B and table S1). The resulting 284 peptide clusters were randomly assigned for unique combinations of quadruple streptavidin-metal (SAv-metal) coding (Fig. 1A and Fig. 7C). This approach avoided false interpretation of the combinatorial pMHC tetramer strategy (29) that would result from T cells expected to cross-react with multiple minor variants of the same peptide. The coded 562-plex pMHC tetramers library was pooled and simultaneously probed on each patient’s peripheral lymphocytes (Fig. 7C) (30). To increase confidence of detection, patient’s cells were evenly divided and independently interrogated by the same 562-plex pMHC tetramers library of two entirely different SAv-metal coding configurations (Fig. 7C-D and S2A). Together with >26 cellular markers (table S2), the signals of 562-plex pMHC tetramers were determined by mass cytometry. A self-validated automatic combinatorial tetramer deconvolution algorithm was used to identify tetramer positive events in an unbiased way (Materials and Methods, Fig. 7D). The correspondences between matching tetramers from two coding configurations were calculated using a bootstrapping statistical analysis (see Materials and Methods, Fig. 1A and Fig. 7D). Finally, the validated antigen-specific CD8<sup>+</sup> T cells for those passing all the deconvolution criteria between two configurations were enumerated, and such approach can be verified by the correlation of control epitopes (e.g. CMV and EBV) detected between the configurations (Fig. 7 and S2A). Using this objective strategy, we were able to detect many unidentified candidate epitopes and their variants in CHB patients, as well as the previous identified epitopes (table S1).

It was hypothesized that the antigen-specific T cell responses could vary across different clinical stages and reflect CHB disease progression. Therefore, this strategy to map potential T cell epitopes across three CHB patient groups (IT, IA and InA) and one group of acutely resolved patients (R) (table S3) was applied. There was no difference in the overall magnitudes of detected antigen-specificities between patient groups (Fig. 9). However, across all tested patients, T cells specific for more epitopes derived from polymerase (P) and core (C) compared to envelope (S) and x (X) proteins were detected, including four epitopes with the highest frequencies observed (Fig. 1B). These were HBV-P-282 (cluster 090, 4

peptide variants), HBV-P-387 (cluster 106, 1 unique peptide), HBV-C-169 (cluster 178, 7 peptide variants), and HBV-C-195v2 (cluster 283, 1 unique peptide) (Fig. 1B-C, Fig. 8 and table S1).

5 Several experiments were performed to validate and assess the HBV relevance of these four epitopes. For three out of four of these epitopes, antigen-specific T cells were detected in some healthy donor (HD) or cord blood (CB) samples (Fig. 1D), but not in HLA-mismatched patients (Fig. 10). The detection of these cells was further reproduced and confirmed by fluorescence flow cytometry pMHC tetramer staining, which showed consistent results (Fig. 10 11A-B). These T cells could also proliferate upon the stimulation by corresponding viral peptides (Fig. 12), except the seemingly unresponsiveness of cells from IT and HD. Although the results obtained were clear in confirming the specificity of T cells stained with HBV-P-282 and HBV-C-195v2 peptide-loaded pMHC tetramers, these cells displayed a mostly naïve-like phenotype (Fig. 1E). It is likely that T cells specific for these epitopes could be naïve T 15 cells with relatively high precursor frequencies due to the non-random nature of TCR recombination, and were similar to HCV-specific CD8<sup>+</sup> naïve precursors that previously described.

A previously identified epitope, HBV-P-387 was observed with relatively high frequency in 20 CHB and also in half of the HD tested, but to a lesser extent in CB samples. As described further in subsequent sections, the heterogeneous phenotypes of these cells were highly variable between patients (Fig. 1E), suggesting the relevant HBV-specific immune response. Given the high prevalence of HBV in Singapore, where the HD were collected, we postulate that such unexpected detection of HBV-specific T cells could be due to the high coverage of 25 vaccination or subclinical infection without the development of anti-HBc antibody (HBcAb), as reported in HBV-exposed health workers (Fig. 1D and Fig. 11C). This remains to be determined. Lastly, among the validated epitopes, T cells specific for HBV-C-169 were only detected in HBV-infected individuals and elevated in patients with viral control (InA and R) (Fig. 1D) compared to patients with high viral load (IT), where these cells were undetectable. 30 Such HBV-specific T cells displayed phenotypic profiles that differed according to the status of HBV-infection (Fig. 1E). These results prompted us to further evaluate the profiles of HBV<sub>pol387</sub> and HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells.

## 2. High-dimensional phenotypic profiling of HBV-specific CD8<sup>+</sup> T cells

Further analyses were directed at HBV<sub>pol387</sub> and HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells because of their higher degrees of phenotypic heterogeneity observed across patients at various stages of CHB. Although both HBV-P-387 (LVVDFSQFSR) and one of the peptides within the HBV-C-169 (STLPETTVVRR) have been previously reported, the phenotypes of these reactive T cells have not been investigated. Unique to HBV<sub>pol387</sub> and HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells is the higher expression of TIGIT compared to other HBV-specific CD8<sup>+</sup> T cells (Fig. 13), as well as elevated PD-1 expression on HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells (Fig. 13).

Unsupervised high-dimensional *t*-SNE visualization and Phenograph cellular clustering were applied to describe the phenotypes of virus-specific CD8<sup>+</sup> T cells across individuals from one large batch of samples run in parallel (Fig. 2). Based on the expression levels of markers indicative of T cell activation, differentiation, trafficking and inhibitory receptors typically associated with T cell exhaustion, 19 cellular clusters were objectively identified and annotated (Fig. 2A).

From this analysis, remarkable heterogeneity of HBV<sub>pol387</sub>-specific CD8<sup>+</sup> T cells was observed. In many instances, several distinct populations specific for this one epitope could be seen even within individual patients. Despite such diverse phenotypes, this epitope sequence was highly conserved across all patients and time-points subjected to HBV viral sequencing over a decade (Fig. 14, table S4). Quantification of cellular clusters within each T cell antigen-specificity were performed across all batches of experiments using cluster-specific gating strategies (Fig. 15) to test for compositional differences associated with the status of HBV infection. Besides cluster 8 (C8), all of the cellular clusters occupied by HBV<sub>pol387</sub>-specific CD8<sup>+</sup> T cells expressed 2B4 with heterogeneous phenotype indicative of different T cell memory status, whereas only cluster 9 (C9) showed elevated expression of PD-1 (Fig. 2). In terms of relationships with the status of infection, a significant enrichment of cells with C13-like phenotypes (~80%) in IT patients, expressing CXCR3<sup>+</sup>CD27<sup>hi</sup>CD127<sup>hi</sup> (Fig. 2A-B and Fig. 16A-C) was observed. Significant differences were also seen for C8 in IA patients, which was highlighted by the co-expression of CD45RO<sup>hi</sup>CCR4<sup>+</sup>HVEM<sup>+</sup> with a T<sub>CM</sub> phenotype. The

phenotypically similar region C6+C17 was preferentially occupied by InA, which were notably absent from IT patients (Fig. 2B and Fig. 16A-C). Such subset was similar to terminal effector memory RA (TEMRA) but expressed CD127<sup>Int</sup>, perhaps suggesting an ongoing T cell response specific for InA but not IT patients. Given that these cells were less expanded in IT patients (Fig. 12) and expressed several memory-associated markers, it suggests that they were experienced but not fully activated, and perhaps such inactivation was regulated by 2B4 in a PD-1-independent manner (Fig. 2A).

Supplementary Table 4: The frequency of viral mutation on selective epitopes in longitudinal patient cohort across HBeAg-seroconversion.

| Patient group      | Patient ID | Epitope        | Sequence     | Time       |             |            |             |            |             |             |  |
|--------------------|------------|----------------|--------------|------------|-------------|------------|-------------|------------|-------------|-------------|--|
|                    |            |                |              | 4-Jan-1999 | 29-May-2000 | 4-Jun-2001 | 29-Jul-2003 | 3-Feb-2004 | 27-Jul-2005 | 10-Aug-2006 |  |
| Non-seroconverters | C3         | 003_HBV-S-95*  | STNRQSGRG    | 0.9623     | 0.9603      | 0.9601     | 0.9619      | 0.9623     | 0.9607      | 0.9611      |  |
|                    |            | 056_HBV-P-136* | VVNHVQTR     | 0.9704     | 0.9692      | 0.9681     | 0.9666      | 0.9698     | 0.9689      | 0.9677      |  |
|                    |            | 090_HBV-P-282* | KAAYSLSISK   | 0.9448     | 0.9436      | 0.9399     | 0.9396      | 0.9443     | 0.9426      | 0.9442      |  |
|                    |            | 106_HBV-P-387  | LVVDFGDFSR   | 0.9556     | 0.9558      | 0.9542     | 0.9509      | 0.9581     | 0.9558      | 0.9556      |  |
|                    |            | 123_HBV-P-486* | STVSLMLLYK   | 0.9552     | 0.9541      | 0.9537     | 0.9514      | 0.9548     | 0.9521      | 0.9560      |  |
|                    |            | 178_HBV-C-169* | STLPETTVVRR  | 0.9607     | 0.9631      | 0.9616     | 0.9592      | 0.9618     | 0.9588      | 0.9578      |  |
|                    |            | 179_HBV-C-173* | TVVRRRGGRSPR | 0.9570     | 0.9572      | 0.9578     | 0.9556      | 0.9584     | 0.9559      | 0.9563      |  |
|                    |            | 181_HBV-X-80*  | TTVNAHRNLPR  | 0.9474     | 0.9462      | 0.9463     | 0.9413      | 0.9500     | 0.9480      | 0.9433      |  |
|                    |            | 203_HBV-C-195* | RSQSPRRRRSQ  | 0.9568     | 0.9555      | 0.9551     | 0.9517      | 0.9557     | 0.9542      | 0.9550      |  |
|                    | C4         | 003_HBV-S-95*  | STNRQSGRG    | 0.9851     | 0.9805      | 0.9841     | 0.9899      |            |             |             |  |
|                    |            | 056_HBV-P-136* | VVNHVQTR     | 0.9764     | 0.9719      | 0.9419     | 0.8343      |            |             |             |  |
|                    |            | 090_HBV-P-282* | KAAYSLSISK   | 0.9784     | 0.9415      | 0.9656     | 0.9580      |            |             |             |  |
|                    |            | 106_HBV-P-387  | LVVDFGDFSR   | 0.9805     | 0.9573      | 0.9815     | 0.9518      |            |             |             |  |
|                    |            | 123_HBV-P-486* | STVSLMLLYK   | 0.9779     | 0.9460      | 0.9706     | 0.9433      |            |             |             |  |
|                    |            | 178_HBV-C-169* | STLPETTVVRR  | 0.9832     | 0.9622      | 0.9824     | 0.9571      |            |             |             |  |
|                    |            | 179_HBV-C-173* | TVVRRRGGRSPR | 0.9801     | 0.9601      | 0.9807     | 0.9540      |            |             |             |  |
|                    |            | 181_HBV-X-80*  | TTVNAHRNLPR  | 0.9751     | 0.9412      | 0.9716     | 0.9406      |            |             |             |  |
|                    |            | 203_HBV-C-195* | RSQSPRRRRSQ  | 0.9768     | 0.9536      | 0.9749     | 0.9454      |            |             |             |  |
|                    | C5         | 003_HBV-S-95*  | STNRQSGRG    | 0.9839     | 0.9635      | 0.9633     | 0.9632      | 0.9801     |             |             |  |
|                    |            | 056_HBV-P-136* | VVNHVQTR     | 0.9614     | 0.9594      | 0.9636     | 0.9633      | 0.9808     |             |             |  |
|                    |            | 090_HBV-P-282* | KAAYSLSISK   | 0.9488     | 0.9462      | 0.9494     | 0.9470      | 0.9713     |             |             |  |
|                    |            | 106_HBV-P-387  | LVVDFGDFSR   | 0.9617     | 0.9595      | 0.9625     | 0.9638      | 0.9810     |             |             |  |
|                    |            | 123_HBV-P-486* | STVSLMLLYK   | 0.9577     | 0.9579      | 0.9614     | 0.9606      | 0.9819     |             |             |  |
|                    |            | 178_HBV-C-169* | STLPETTVVRR  | 0.9632     | 0.9615      | 0.9646     | 0.9663      | 0.9818     |             |             |  |
|                    |            | 179_HBV-C-173* | TVVRRRGGRSPR | 0.9566     | 0.9597      | 0.9615     | 0.9607      | 0.9786     |             |             |  |
|                    |            | 181_HBV-X-80*  | TTVNAHRNLPR  | 0.9481     | 0.9477      | 0.9546     | 0.9515      | 0.9735     |             |             |  |
|                    |            | 203_HBV-C-195* | RSQSPRRRRSQ  | 0.9611     | 0.9593      | 0.9608     | 0.9616      | 0.9760     |             |             |  |
|                    | C7         | 003_HBV-S-95*  | STNRQSGRG    | 0.9629     | 0.9630      | 0.9594     | 0.9610      | 0.9643     | 0.9624      | 0.9623      |  |
|                    |            | 056_HBV-P-136* | VVNHVQTR     | 0.9593     | 0.9600      | 0.9592     | 0.9564      | 0.9595     | 0.9586      | 0.9587      |  |
|                    |            | 090_HBV-P-282* | KAAYSLSISK   | 0.9454     | 0.9458      | 0.9412     | 0.9436      | 0.9471     | 0.9430      | 0.9417      |  |
|                    |            | 106_HBV-P-387  | LVVDFGDFSR   | 0.9558     | 0.9569      | 0.9530     | 0.9517      | 0.9587     | 0.9564      | 0.9566      |  |
|                    |            | 123_HBV-P-486* | STVSLMLLYK   | 0.9494     | 0.9513      | 0.9473     | 0.9490      | 0.9517     | 0.9489      | 0.9484      |  |
|                    |            | 178_HBV-C-169* | STLPETTVVRR  | 0.9595     | 0.9619      | 0.9597     | 0.9610      | 0.9611     | 0.9591      | 0.9595      |  |
|                    |            | 179_HBV-C-173* | TVVRRRGGRSPR | 0.9570     | 0.9579      | 0.9579     | 0.9578      | 0.9589     | 0.9568      | 0.9529      |  |
|                    |            | 181_HBV-X-80*  | TTVNAHRNLPR  | 0.9430     | 0.9449      | 0.9405     | 0.9398      | 0.9484     | 0.9436      | 0.9425      |  |
|                    |            | 203_HBV-C-195* | RSQSPRRRRSQ  | 0.9544     | 0.9553      | 0.9570     | 0.9525      | 0.9552     | 0.9521      | 0.9493      |  |

Supplementary Table 4, Continued.

| Patient group       | Patient ID | Epitope        | Sequence     | Time        |             |             |             |            |            |
|---------------------|------------|----------------|--------------|-------------|-------------|-------------|-------------|------------|------------|
|                     |            |                |              | 17-Aug-1998 | 26-Oct-1998 | 17-Jan-1999 | 14-Mar-2001 | 3-Jun-2001 | 9-Feb-2004 |
| HBeAg-seroconverter | S1         | 003_HBV-S-95*  | STNRGGGGRG   | 0.9799      | 0.9815      | 0.9812      | 0.9852      | 0.9864     | 0.9770     |
|                     |            | 056_HBV-P-136* | VVNHVFGTR    | 0.9854      | 0.9862      | 0.9853      | 0.9874      | 0.9826     | 0.9747     |
|                     |            |                | VVNHVFGTR    |             |             |             |             |            | 0.9809     |
|                     |            | 090_HBV-P-282* | KAAYSLISTSK  | 0.9748      | 0.9725      | 0.9758      | 0.9709      | 0.9428     | 0.9697     |
|                     |            | 106_HBV-P-387  | LVVDFSGFSR   | 0.9822      | 0.9821      | 0.9816      | 0.9806      | 0.9526     | 0.9801     |
|                     |            | 120_HBV-P-486* | VVSLMLLYK    | 0.9819      | 0.9810      | 0.9818      | 0.9547      | 0.9485     | 0.9784     |
|                     |            | 178_HBV-C-169* | STLPETTIVRR  | 0.9825      | 0.9827      | 0.9810      | 0.9828      | 0.9590     | 0.9870     |
|                     |            | 179_HBV-C-175* | TVVRRHGRSPR  | 0.9793      | 0.9804      | 0.9798      | 0.9795      | 0.7718     | 0.7836     |
|                     |            |                | TVVRRHGRSPR  |             |             |             |             | 0.7843     | 0.6838     |
|                     |            | 183_HBV-S-80*  | TTVNAHRNLPK  |             |             |             |             |            |            |
|                     |            |                | TTVNAHRNLPK  | 0.9731      | 0.9726      | 0.9751      | 0.9484      | 0.9382     | 0.9749     |
|                     |            |                | TTVNAHGRNLPK |             |             |             | 0.9138      |            |            |
|                     |            |                | TTVNAHGRNLPK |             |             |             | 0.9538      |            |            |
|                     |            | 283_HBV-C-195* | RSQSPRRRSQ   | 0.9772      | 0.9786      | 0.9777      | 0.9538      | 0.9565     | 0.9779     |
| S2                  |            | 003_HBV-S-95*  | STNRGGGGRG   | 0.9824      | 0.9794      | 0.9809      | 0.9884      | 0.9878     | 0.9818     |
|                     |            | 056_HBV-P-136* | VVNHVFGTR    | 0.9888      | 0.9835      | 0.9885      | 0.9831      | 0.9884     | 0.9811     |
|                     |            |                | VVNHVFGTR    |             |             |             |             | 0.9335     | 0.9623     |
|                     |            |                | VVNHVFGTR    |             |             |             |             | 0.9879     | 0.9558     |
|                     |            |                | VVNHVFGTR    |             |             |             |             | 0.9247     |            |
|                     |            |                | VVNHVFGTR    |             |             |             |             |            | 0.9817     |
|                     |            | 090_HBV-P-282* | KAAYSLISTSK  | 0.9474      | 0.9713      | 0.9477      | 0.9727      | 0.9516     | 0.9700     |
|                     |            |                | KAAYSLISTSK  |             |             |             |             | 0.9700     | 0.9880     |
|                     |            |                | KAAYSLISTSK  |             |             |             |             |            | 0.9500     |
|                     |            | 106_HBV-P-387  | LVVDFSGFSR   | 0.9613      | 0.9818      | 0.9620      | 0.9821      | 0.9784     | 0.9800     |
|                     |            | 120_HBV-P-486* | VVSLMLLYK    | 0.9573      | 0.9827      | 0.9818      | 0.9839      | 0.9801     | 0.9844     |
|                     |            |                | VVSLMLLYK    |             |             |             |             |            | 0.9700     |
|                     |            | 178_HBV-C-169* | STLPETTIVRR  | 0.9640      | 0.9822      | 0.9667      | 0.9821      | 0.9700     | 0.9700     |
|                     |            |                | STLPETTIVRR  |             |             |             |             |            | 0.9700     |
|                     |            |                | STLPETTIVRR  |             |             |             |             |            | 0.9700     |
|                     |            |                | STLPETTIVRR  |             |             |             |             |            | 0.9700     |
|                     |            | 179_HBV-C-175* | TVVRRHGRSPR  | 0.9853      | 0.9794      | 0.9809      | 0.9808      | 0.9568     | 0.9873     |
|                     |            |                | TVVRRHGRSPR  |             |             | 0.9808      | 0.9808      | 0.9478     | 0.9758     |
|                     |            |                | TVVRRHGRSPR  |             |             |             |             |            | 0.9758     |
|                     |            |                | TVVRRHGRSPR  |             |             |             |             |            | 0.9758     |
|                     |            | 183_HBV-S-80*  | TTVNAHRNLPK  | 0.9772      | 0.9429      | 0.9310      | 0.9712      | 0.9456     | 0.9831     |
|                     |            |                | TTVNAHRNLPK  |             |             |             |             |            | 0.9800     |
|                     |            |                | TTVNAHGRNLPK |             |             |             |             |            | 0.9800     |
|                     |            | 283_HBV-C-195* | RSQSPRRRSQ   | 0.9589      | 0.9780      | 0.9639      | 0.9788      | 0.9804     | 0.9809     |
| S3                  |            | 003_HBV-S-95*  | STNRGGGGRG   | 0.9864      | 0.9418      | 0.9818      |             |            |            |
|                     |            |                | STNRGGGGRG   |             | 0.9331      |             |             |            |            |
|                     |            | 056_HBV-P-136* | VVNHVFGTR    | 0.9200      | 0.9098      | 0.9752      |             |            |            |
|                     |            |                | VVNHVFGTR    | 0.9160      |             |             |             |            |            |
|                     |            |                | VVNHVFGTR    | 0.9327      | 0.9735      | 0.9814      |             |            |            |
|                     |            | 090_HBV-P-282* | KAAYSLISTSK  | 0.9830      | 0.9118      | 0.9785      |             |            |            |
|                     |            |                | KAAYSLISTSK  | 0.9588      |             |             |             |            |            |
|                     |            |                | KAAYSLISTSK  |             | 0.9102      |             |             |            |            |
|                     |            |                | KAAYSLISTSK  |             | 0.9327      |             |             |            |            |
|                     |            |                | KAAYSLISTSK  |             | 0.9336      |             |             |            |            |
|                     |            | 106_HBV-P-387  | LVVDFSGFSR   | 0.9833      | 0.9820      | 0.9800      |             |            |            |
|                     |            | 120_HBV-P-486* | VVSLMLLYK    | 0.9413      | 0.9841      | 0.9812      |             |            |            |
|                     |            |                | VVSLMLLYK    | 0.9833      |             |             |             |            |            |
|                     |            |                | VVSLMLLYK    | 0.9833      |             |             |             |            |            |
|                     |            | 178_HBV-C-169* | STLPETTIVRR  | 0.9817      | 0.9810      | 0.9794      |             |            |            |
|                     |            | 179_HBV-C-175* | TVVRRHGRSPR  | 0.9807      | 0.9806      | 0.9773      |             |            |            |
|                     |            | 183_HBV-S-80*  | TTVNAHRNLPK  | 0.9138      | 0.7518      | 0.9617      |             |            |            |
|                     |            |                | TTVNAHRNLPK  | 0.9662      |             |             |             |            |            |
|                     |            |                | TTVNAHRNLPK  |             | 0.9442      | 0.9737      |             |            |            |
|                     |            |                | TTVNAHGRNLPK |             | 0.9572      |             |             |            |            |
|                     |            | 283_HBV-C-195* | RSQSPRRRSQ   | 0.9838      | 0.6616      | 0.9712      |             |            |            |
|                     |            |                | RSQSPRRRSQ   |             | 0.9779      | 0.9738      |             |            |            |
|                     |            |                | RSQSPRRRSQ   |             | 0.9834      |             |             |            |            |

Table S4 – continued

|                     |              | 4-Jan-1992     |              |        |        | 16-Feb-1993 |        | 20-Jun-1994 |        | 11-Jan-1995 |  |  |  |
|---------------------|--------------|----------------|--------------|--------|--------|-------------|--------|-------------|--------|-------------|--|--|--|
| HBeAg-seroconverter | S4           | 003_HBV-S-95*  | STNRQSGRQ    | 0.9635 | 0.9588 | 0.9654      | 0.9790 |             |        |             |  |  |  |
|                     |              | 056_HBV-P-136* | VVNHYFQTR    | 0.9632 | 0.9680 | 0.9645      | 0.9796 |             |        |             |  |  |  |
|                     |              | 090_HBV-P-283* | KAAYSLISTSK  | 0.9413 | 0.9373 | 0.9454      | 0.9552 |             |        |             |  |  |  |
|                     |              |                | KAAYSLISTSK  |        |        |             | 0.9581 |             |        |             |  |  |  |
|                     |              | 106_HBV-P-387  | LVVDFSGFSR   | 0.9528 | 0.9587 | 0.9555      | 0.9819 |             |        |             |  |  |  |
|                     |              | 125_HBV-P-486* | YVSLMLLYK    | 0.9539 | 0.9472 | 0.9530      | 0.9815 |             |        |             |  |  |  |
|                     |              | 178_HBV-C-169* | STLPETTIVRR  | 0.9614 | 0.9616 | 0.9596      | 0.9805 |             |        |             |  |  |  |
|                     |              | 179_HBV-C-175* | TVVRRRGGRSPR | 0.9585 | 0.9569 | 0.9545      | 0.9763 |             |        |             |  |  |  |
|                     |              | 183_HBV-X-80*  | TTVNAHRNLPK  | 0.9395 | 0.9425 | 0.9397      | 0.9580 |             |        |             |  |  |  |
|                     |              |                | TTVNAHRNLPK  |        |        |             | 0.9580 |             |        |             |  |  |  |
|                     |              | 283_HBV-C-195* | RSQSPRRRSQ   | 0.9607 | 0.9588 | 0.9589      | 0.9794 |             |        |             |  |  |  |
|                     |              |                |              |        |        |             |        |             |        |             |  |  |  |
|                     |              | 14-Mar-1991    |              |        |        |             |        |             |        |             |  |  |  |
| 3-Apr-1993          |              |                |              |        |        |             |        |             |        |             |  |  |  |
|                     | S5           | 003_HBV-S-95*  | STNRQSGRQ    | 0.9813 | 0.9594 |             |        |             |        |             |  |  |  |
|                     |              | 056_HBV-P-136* | VVNHYFQTR    | 0.8592 | 0.9603 |             |        |             |        |             |  |  |  |
|                     |              |                | VVDHYFQTR    | 0.0417 |        |             |        |             |        |             |  |  |  |
|                     |              |                | VVNHYFQTR    | 0.0783 |        |             |        |             |        |             |  |  |  |
|                     |              | 090_HBV-P-283* | KAAYSLISTSK  | 0.9749 | 0.9080 |             |        |             |        |             |  |  |  |
|                     |              |                | KAAYSLISTSK  |        | 0.9087 |             |        |             |        |             |  |  |  |
|                     |              | 106_HBV-P-387  | LVVDFSGFSR   | 0.9823 | 0.9586 |             |        |             |        |             |  |  |  |
|                     |              | 125_HBV-P-486* | YVSLMLLYK    | 0.9822 | 0.9568 |             |        |             |        |             |  |  |  |
|                     |              | 178_HBV-C-169* | STLPETTIVRR  | 0.9833 | 0.9611 |             |        |             |        |             |  |  |  |
|                     |              | 179_HBV-C-175* | TVVRRRGGRSPR | 0.9463 | 0.9413 |             |        |             |        |             |  |  |  |
|                     |              |                | TVVRRRGGRSPR | 0.0336 | 0.0301 |             |        |             |        |             |  |  |  |
|                     |              | 183_HBV-X-80*  | TTVNAHRNLPK  | 0.0289 | 0.0301 |             |        |             |        |             |  |  |  |
|                     |              |                | TTVNAHRNLPK  | 0.0308 | 0.0304 |             |        |             |        |             |  |  |  |
|                     | TTVNAHRNLPK  |                | 0.0304       |        |        |             |        |             |        |             |  |  |  |
|                     | TTVNAHRNLPK  |                | 0.0304       |        |        |             |        |             |        |             |  |  |  |
| 283_HBV-C-195*      | RSQSPRRRSQ   | 0.9755         | 0.9578       |        |        |             |        |             |        |             |  |  |  |
|                     |              |                |              |        |        |             |        |             |        |             |  |  |  |
| 20-Oct-1991         |              |                |              |        |        |             |        |             |        |             |  |  |  |
| 9-Dec-1991          |              |                |              |        |        |             |        |             |        |             |  |  |  |
| 4-Aug-1992          |              |                |              |        |        |             |        |             |        |             |  |  |  |
| 23-Feb-1994         |              |                |              |        |        |             |        |             |        |             |  |  |  |
| 6-Jun-1996          |              |                |              |        |        |             |        |             |        |             |  |  |  |
| 10-Apr-2001         |              |                |              |        |        |             |        |             |        |             |  |  |  |
|                     | S6           | 003_HBV-S-95*  | STNRQSGRQ    | 0.9678 | 0.9796 | 0.9793      | 0.9792 | 0.9799      | 0.9825 |             |  |  |  |
|                     |              | 056_HBV-P-136* | VVNHYFQTR    | 0.9617 | 0.9836 | 0.9848      | 0.9871 | 0.9775      | 0.9770 |             |  |  |  |
|                     |              |                | VVNHYFQTR    | 0.0133 |        |             |        |             |        |             |  |  |  |
|                     |              | 090_HBV-P-283* | KAAYSLISTSK  | 0.9547 | 0.9627 | 0.9565      | 0.9639 | 0.9594      | 0.7289 |             |  |  |  |
|                     |              |                | KAAYSLISTSK  | 0.0137 |        | 0.0318      |        |             | 0.1834 |             |  |  |  |
|                     |              |                | KAAYSLISTSK  |        |        |             |        |             | 0.0847 |             |  |  |  |
|                     |              | 106_HBV-P-387  | LVVDFSGFSR   | 0.9810 | 0.9826 | 0.9819      | 0.9819 | 0.9818      | 0.9794 |             |  |  |  |
|                     |              | 125_HBV-P-486* | YVSLMLLYK    | 0.9848 | 0.9875 | 0.9838      | 0.9719 | 0.9811      | 0.9770 |             |  |  |  |
|                     |              | 178_HBV-C-169* | STLPETTIVRR  | 0.9830 | 0.9806 | 0.9832      | 0.9827 | 0.9812      | 0.9663 |             |  |  |  |
|                     |              |                | STLPETTIVRR  |        |        |             |        |             | 0.9829 |             |  |  |  |
|                     |              | 179_HBV-C-175* | TVVRRRGGRSPR | 0.9794 | 0.9784 | 0.9779      | 0.9807 | 0.9789      | 0.9627 |             |  |  |  |
|                     |              |                | TVVRRRGGRSPR |        |        |             |        |             | 0.9663 |             |  |  |  |
|                     |              | 183_HBV-X-80*  | TTVNAHRNLPK  |        |        |             | 0.9838 |             |        |             |  |  |  |
|                     | TTVNAHRNLPK  | 0.9807         | 0.9753       | 0.9788 | 0.9794 | 0.9716      | 0.0872 |             |        |             |  |  |  |
|                     | TTVNAHRNLPK  |                |              |        |        |             | 0.9843 |             |        |             |  |  |  |
| 283_HBV-C-195*      | RSQSPRRRSQ   | 0.9815         | 0.9783       | 0.9760 | 0.9816 | 0.9113      | 0.2492 |             |        |             |  |  |  |
|                     | RTQSPRRRSQ   |                |              |        |        | 0.9848      | 0.7289 |             |        |             |  |  |  |
|                     |              |                |              |        |        |             |        |             |        |             |  |  |  |
| 19-Sep-1990         |              |                |              |        |        |             |        |             |        |             |  |  |  |
| 29-Oct-1991         |              |                |              |        |        |             |        |             |        |             |  |  |  |
| 15-Sep-1992         |              |                |              |        |        |             |        |             |        |             |  |  |  |
| 19-Oct-1993         |              |                |              |        |        |             |        |             |        |             |  |  |  |
|                     | S7           | 003_HBV-S-95*  | STNRQSGRQ    | 0.9645 | 0.9831 | 0.9841      | 0.9832 |             |        |             |  |  |  |
|                     |              | 056_HBV-P-136* | VVNHYFQTR    | 0.7695 | 0.6757 | 0.9503      | 0.9849 |             |        |             |  |  |  |
|                     |              |                | VVDHYFQTR    | 0.0609 | 0.1136 | 0.0343      |        |             |        |             |  |  |  |
|                     |              |                | VVNHYFQTR    | 0.1436 | 0.1097 |             |        |             |        |             |  |  |  |
|                     |              | 090_HBV-P-283* | KAAYSLISTSK  | 0.9152 | 0.8284 | 0.9782      | 0.9792 |             |        |             |  |  |  |
|                     |              |                | KAAYSLISTSK  | 0.0363 | 0.1230 |             |        |             |        |             |  |  |  |
|                     |              |                | KAAYSLISTSK  |        | 0.1221 |             |        |             |        |             |  |  |  |
|                     |              | 106_HBV-P-387  | LVVDFSGFSR   | 0.9817 | 0.9826 | 0.9819      | 0.9850 |             |        |             |  |  |  |
|                     |              | 125_HBV-P-486* | YVSLMLLYK    | 0.9793 | 0.9414 | 0.9685      | 0.9848 |             |        |             |  |  |  |
|                     |              |                | YVSLMLLYK    |        | 0.9413 | 0.9237      |        |             |        |             |  |  |  |
|                     |              | 178_HBV-C-169* | STLPETTIVRR  | 0.9798 | 0.9766 | 0.9832      | 0.9846 |             |        |             |  |  |  |
|                     |              |                | STLPETTIVRR  |        | 0.9799 |             |        |             |        |             |  |  |  |
|                     |              | 179_HBV-C-175* | TVVRRRGGRSPR | 0.9277 | 0.9381 | 0.9786      | 0.9808 |             |        |             |  |  |  |
|                     | TVVRRRGGRSPR | 0.0568         | 0.0486       |        |        |             |        |             |        |             |  |  |  |
| 183_HBV-X-80*       | TTVNAHRNLPK  |                |              | 0.9787 | 0.9438 |             |        |             |        |             |  |  |  |
|                     | TTVNAHRNLPK  | 0.7742         | 0.8143       | 0.9794 | 0.9318 |             |        |             |        |             |  |  |  |
|                     | TTVNAHRNLPK  | 0.7997         | 0.7962       |        |        |             |        |             |        |             |  |  |  |
|                     | TTVNAHRNLPK  |                | 0.8134       |        |        |             |        |             |        |             |  |  |  |
| 283_HBV-C-195*      | RSQSPRRRSQ   | 0.9733         | 0.9484       | 0.9814 | 0.9799 |             |        |             |        |             |  |  |  |
|                     | RSQSPRRRSQ   |                | 0.9486       |        |        |             |        |             |        |             |  |  |  |

Table S4 – continued

Comprised of seven peptide-sequence variants (Fig. 1C and table S1), HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells had significantly greater frequencies (Fig. 1D) and pMHC tetramer staining intensity in patients with viral control (InA and R) (Fig. 16D). No T cells specific for this epitope were detected in IT patients or HD (Fig. 1D), except one IT patient had detectable cells (0.00193%) just below the imposed cut-off frequency (0.002%) (Fig. 16E). *t*-SNE and Phenograph analysis showed that C8 and C11 were significantly enriched in InA and R, whereas C1 and C4 tended to be more prevalent in IA and InA patients. This is in-line with hierarchical clustering (Fig. 1E) that segregated individuals into different clinical stages based on the phenotypes of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells (Fig. 1E), which was not observed for EBV-specific CD8<sup>+</sup> T cells analyzed in parallel (Fig. 16F). In contrast to the elevated expression of CD57, PD-1 and TIGIT seen for these cells in IA patients, HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells from InA and R patients significantly expressed CD27, CD28, CD45RO, CD127 and CXCR3 (Fig. 2D-E), suggesting they were long-lived memory T cells that were associated with a high degree of viral control. However, it is interesting to note that these cells from R patients expressed relatively high levels of PD-1 and TIGIT despite their presumed clearance of virus. Albeit, these cells derived from R patients differed in the expression of other memory-associated markers and CD57 compared to IA patients. Importantly, the PD-1-expressing HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells from R patients exhibited high level of IL-7R (CD127), indicating they were not T<sub>EX</sub>, but likely to be long-lived memory cells. It is plausible that such PD-1 expression denoted a sign of activation, or adaptation rather than exhaustion, and was induced by strong TCR-viral antigen engagement (Fig. 16D). It is also possible that the R patients were not completely cleared of the virus, and these PD-1-expressing HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells still actively inhibit the virus. In contrast, HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells from InA patients displayed intermediate levels of CD127 and CXCR3, and similar levels of CD27, CD28 and CD45RO compared to R patients. These cells in InA also had diminished expression of PD-1 compared to other groups (Fig. 2D-E). This can be explained by the low-to-undetectable viral load during the stage of InA, leading to lesser TCR-viral antigen engagement compared to IA patients who had high viral load.

Lastly, Scorpius ( R. Cannoodt *et al.*, SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. *bioRxiv*), a trajectory inference method, was

applied to compute the trajectory of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells across three clinical stages using the patient-wise expression of eight phenotypic markers that showed statistical significances (Fig. 2F). Our analysis indicated that decreased expression of PD-1, TIGIT and CD57 together with the increased expression of CD27, CD28, CXCR3, CD45RO and CD127 were associated with the inferred status of infection. Furthermore, patients associated with viral control (InA and R) were nicely separated and toward the end of the trajectory, whereas IA patients were on the opposite side. Thus, it was demonstrated the highly heterogeneous HBV-specific CD8<sup>+</sup> T cells during the progression of CHB using several high-dimensional analytical approaches, and such multifactorial cellular responses targeting HBV<sub>pol387</sub> and HBV<sub>core169</sub> were able to delineate patients into their clinical stages.

### 3. Multifactorial interrelations of inhibitory receptors on virus-specific T cells

Next, the relationships between various categories of cellular markers expressed by each of the antigen-specific T cells analyzed, with a special focus on nine different inhibitory receptors was evaluated. To directly assess these relationships, One-Dimensional Soli-Expression by Nonlinear Stochastic Embedding (One-SENSE) was employed. One-SENSE works by reducing dimensionality of each category of markers into a one-dimensional *t*-SNE map that can be plotted in conjunction with alternative categories of markers mapped into additional one-dimensional *t*-SNE maps. In this way, cells are separately arranged based on their categorical expression and then relationships between the categories can be intuitively visualized and described.

In this case, the categories assigned (table S2) were: “Differentiation+TNFR” (markers of differentiation and tumor necrosis factor receptor superfamily), “Inhibitory” (inhibitory receptors) and “Trafficking” (chemokine receptors). The three derived axes called out the cellular subsets objectively with all possible protein co-expressions (Fig. 3A). The profiles of T cells specific for four different epitopes, including HBV<sub>pol282</sub>, HBV<sub>pol387</sub> and HBV<sub>core169</sub>, and one that derived from EBV (EBV<sub>EBNA3B</sub>) were plotted and compared. In general, of the observed combinations of inhibitory receptors expressed by the antigen-specific T cells across patients, one subset displayed HVEM<sup>int</sup>2B4<sup>+</sup>TIGIT<sup>+</sup>CD160<sup>+</sup>PD-1<sup>lo</sup>, and was mainly contributed by subpopulations of HBV<sub>pol387</sub> and EBV<sub>EBNA3B</sub>-specific CD8<sup>+</sup> T cells. One-SENSE

analysis also showed that 2B4 and HVEM<sup>lo</sup> were expressed by most of the antigen-specific T cells, but limited expressions of LAG-3, TIM-3 and CTLA-4 (Fig. 3A and Fig. 13). Additionally, it was found that the majority of PD-1<sup>+</sup> cells did not express 2B4 and TIGIT but not CD160.

5 By plotting the cells based on the patient groups, it was noted that the phenotypes of HBV<sub>pol387</sub> and HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells (blue and red) were most heavily influenced by status of infection the distinguishing features of these highly diverse cells are labeled. This heterogeneity can be best presented by HBV<sub>pol387</sub> and HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells. HBV<sub>pol387</sub> from IT had a relatively homogeneous phenotype of in terms of memory-  
 10 associated and trafficking receptors, but varied in four distinct co-expressions of inhibitory receptors (Fig. 3A). In other clinical stages, greater diversity was observed for HBV<sub>pol387</sub>-specific CD8<sup>+</sup> T cells in terms of memory versus effector-associated markers and these also had complex relationships with the patterns of inhibitory receptors co-expressed (Fig. 3A). Similar examination of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells was somewhat limited due to the low  
 15 cell numbers but also showed a greater degree of heterogeneity than might be expected. Nonetheless, this representation was consistent with that described above (Fig. 2).

The numbers of co-expressed inhibitory receptors on these cells using a Boolean strategy (Fig. 3B-C) were further quantified. In line with One-SENSE analysis, no HBV-specific cellular  
 20 subsets that accumulated all inhibitory receptors (Fig. 3A-C and Fig. 11) were detected. Although the different co-expressions of inhibitory receptors on HBV<sub>pol387</sub>-specific CD8<sup>+</sup> T cells can be shown by One-SENSE, there were no differences in terms of the numbers accumulated on the cells across patient groups. Conversely, IA patients had significant higher numbers of inhibitory receptors on HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells compared to  
 25 patients with viral control (Fig. 3C). Together with the visualization of One-SENSE, our data suggested a highly heterogeneous antigen-specific phenotype rather than a simple accumulation of so-called "Exhaustion markers" during CHB, and the diverse co-expressions of inhibitory receptors had different relationships with cellular differentiation and trafficking profiles that may associate with disease stages.

30

5. Functional capacity of HBV-specific CD8<sup>+</sup> T cells

To address the relationships of functional capacity and inhibitory receptors, patient's cells were pulsed and expanded using short-term *in vitro* peptide stimulation and then assessed their functional responses using intracellular cytokine staining (Fig. 12). One-SENSE analysis was used to delineate virus-specific CD8<sup>+</sup> T cells into five major heterogeneous functional subsets on the basis of the relationships between categories of "Functions", "Inhibitory" and "Differentiation+TNFR" (Fig. 4A and table S2). For each category (axes of One-SENSE plots), all possible cellular subsets were described by heatplots and descriptive labels. Biaxial plots of the most relevant markers for these five functionally distinct subsets were also represented (Fig. 4B), and the relative composition of these subsets was quantified (Fig. 4C). Overall, this analysis highlights nonlinear relationships between inhibitory receptors and functional capacity on virus-specific CD8<sup>+</sup> T cells upon antigen recall in CHB.

Regardless of patient groups, a multi-functional subset was present (Fig. 4A, green box) and mainly contributed by HBV<sub>pol387</sub> and HBV<sub>env304</sub>-specific CD8<sup>+</sup> T cells expressing MIP-1β<sup>+</sup>GrzA<sup>+</sup>GrzK<sup>lo</sup>Perforin<sup>+</sup> but CD107a<sup>lo</sup> (Fig. 4A-B and Fig. 18A). Another population of these cells (blue box) with otherwise similar phenotypic profiles were distinctly less able to produce cytokines but were GrzA<sup>+</sup>GrzK<sup>+</sup>Perforin<sup>int</sup> without the degranulation marker CD107a, suggesting an alternative form of dysfunctional T cells that was associated with the expression of 2B4 and TIGIT but not PD-1 (Fig. 4A-B and Fig. 18B). These cells were mostly composed of HBV<sub>env304</sub>-specific CD8<sup>+</sup> T cells from CHB patients (Fig. 4A and Fig. 18A), which had expanded greatly in response to *in vitro* peptide stimulation (Fig. 12). Unlike the other inhibitory receptors, CD160 and HVEM were largely reduced on cells producing effector functions upon TCR stimulation. The sustained HVEM was mostly expressed by the non-functional (black box) subset of HBV<sub>pol282</sub>-specific cells (Fig. 4A-B and Fig. 18A-B). Such naïve-like and unresponsive T cells present a different type of dysfunctional T cells, perhaps associated with their expression of BTLA and CD160 prior to antigen recall (Fig. 13). It was also found that a fraction of HBV<sub>env304</sub>-specific CD8<sup>+</sup> T cells from R patients had similar functional profiles as EBV and IAV-specific CD8<sup>+</sup> T cells that were PD-1<sup>+</sup>LAG-3<sup>+</sup>TIM-3<sup>lo</sup>, expressing IFN-γ<sup>+</sup>TNF-α<sup>hi</sup>MIP-1β<sup>hi</sup>GM-CSF<sup>hi</sup> (yellow box) (Fig. 4A and Fig. 18A and C). Moreover, HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells were in the unique pluri-functional (red box) subset co-producing various non-cytolytic and cell recruiting factors (GrzA<sup>+</sup>GrzK<sup>+</sup>IFN-γ<sup>+</sup>TNF-α<sup>lo</sup>MIP-1β<sup>+</sup>GM-CSF<sup>int</sup>CD107a<sup>+</sup>) despite reciprocally co-expressed five inhibitory receptors

including PD-1 (Fig. 4A-B). Interestingly, this subset exhibited high levels of TNFR costimulatory receptors (OX40, GITR, 4-1BB and CD27), suggesting the greater activation and memory status. In this analysis, the major differences observed between patient groups were in the profiles of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells. Patients with better viral control (R > 5 InA > IA) displayed significantly higher the frequencies of pluri-functional subset of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells. In contrast, the opposite trend was observed for the frequencies of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells within the MIP1-β<sup>+</sup> multi-functional subset (Fig. 4C).

10 Together with the abovementioned data, it was concluded that the immune responses of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells were linked to viral control. The analysis also displays complex orchestrations rather than a simple linear relationship between inhibitory receptors and functional capacity on virus-specific CD8<sup>+</sup> T cells during CHB.

## 15 6. Clinical stage-dependent landscapes of virus-specific TCR

How T cell receptors are selected over the course of CHB is largely unknown. Hence, the pMHC tetramer-stained cells (Fig. 11B) were sorted and sequenced the β chain of epitope-specific TCRs (HBV<sub>pol282</sub>, HBV<sub>pol387</sub>, HBV<sub>core169</sub> and HBV<sub>core195</sub>) across various clinical stages. To 20 map the TCR landscapes of these epitopes, *TCRdist* was applied, an algorithm that generates a distance matrix to quantify and obtain the relative motif similarity of TCR based on their sequences of amino acid (P. Dash *et al.*, Quantifiable predictive features define epitope-specific T cell receptor repertoires). *TCRdist* distance matrix was used to cluster similar TCRs using the unsupervised Phenograph clustering algorithm (J. H. Levine *et al.*, Data-Driven 25 Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis) and then visualized by the *t*-SNE dimensionality reduction algorithm (Fig. 5A). Thereafter, the sequence motifs that formed the basis of each TCR-sequence cluster can be presented (Fig. 5B and Fig. 19A), and the composition of these clusters in terms of the antigen-specificity of each sequence derived from patients can be quantified (Fig. 5C). It was found 30 that TCR cluster 15 and 27 were significantly increased in HBV<sub>pol282</sub>-specific TCRs compared with other epitopes (Fig. 5C-D), suggesting that these motifs were important determinants for recognition of this HBV epitope, and may associate with naïve-like phenotypes. In

addition, the relative usage of each of these TCR-sequence clusters differed significantly between patients grouped by status of HBV infection. Dominated by TRBV5-6, TCR cluster 15 could be observed in all patients except for IA patients, which were instead enriched for the TRBV3-2<sup>+</sup> TCR cluster 12. Interestingly, TCR cluster 27 was solely joined by TRBJ2-6 with highly conserved CDR3 (Fig. 19A). Clusters of TCR-sequence usages were more diverse for T cells specific for HBV<sub>core169</sub> and HBV<sub>pol387</sub>. Nonetheless, HBV<sub>core169</sub>-specific TCR sequences also differed between patient groups. The TRBV3-2<sup>+</sup> TCR cluster 12 (also enriched in HBV<sub>pol282</sub>-specific TCRs) was similarly enriched in IA patients within HBV<sub>core169</sub>-specific TCRs. In addition, HBV<sub>core169</sub>-specific TCRs of R and InA had significant higher usages of TRBV6-6<sup>+</sup> cluster 5 and TRBV28<sup>+</sup> cluster 9, respectively (Fig 5C).

Repertoire diversity and density for each TCR sequence from various cell populations using a TCR diversity metric (TCRdiv) was calculated (P. Dash *et al.*, Quantifiable predictive features define epitope-specific T cell receptor repertoires). Various patterns of this measure were observed among the epitope and patient groups (Fig. 5E). Of note, the TCRdiv scores for HBV<sub>pol387</sub>-specific TCRs have striking similar pattern with the proportion of cellular cluster 13 in HBV<sub>pol387</sub>-specific CD8<sup>+</sup> T cells across patient groups (Fig. 2B), which may be attributed to their relative enrichment in IT and HD groups. Importantly, both epitope-specific and bulk TCRs from HD had overall greater diversity (Fig. 19B), which separated them from CHB patients at the molecular level. In contrast to the uniformity of total CD8<sup>+</sup> T cells (Fig. 19C), CDR3 lengths were skewed between epitopes and patient groups. Overall, these findings fit with the previously described phenotypic differences observed in different patient groups. Thus, our analysis indicated the biased TCR repertoire usages during CHB were epitope and clinical stage-dependent.

## 7. Shared HBV<sub>core169</sub>-specific TCR clones in patients with viral control

Focusing on the HBV<sub>core169</sub>-specific response and inquiring a recently curated TCR database, several previously unidentified public HBV<sub>core169</sub>-specific TCR $\beta$  clones that were shared between individuals in a clinical stage-dependent manner were discovered (Fig. 5F). In particular, a general public clone CASGDSNSPLHF was within the top three TCR clones in all three InA patients tested. Two other special public clones CASSGGQIVYEQYF and

CSARGGRGGDYTF were each identified in two InA patients. An additional special public clone CASSQDWTEAFF was found at low frequencies in two acute resolved patients. That these public TCR clones were not shared across patient groups, further highlights differences in the qualities of T cell responses that occur in acute versus chronic viral infection. Failure to detect public clones in IA patients suggested that the presence of public TCRs were essential for HBV viral control. By using PCA to combine the characteristics of HBV<sub>pol387</sub> and HBV<sub>core169</sub>-specific TCR repertoire and cellular response in the same donors, patient's clinical status can be delineated (Fig. 5G). Lastly, it was found that the observed frequency of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells inversely correlated with their TCR repertoire diversity (Fig. 5H), suggesting the selective expansion of T cell clonotypes after viral clearance.

#### 8. Phenotypic dynamic of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells

To assess the characteristic changes of antigen-specific T cells over the course of infection, selected HBV epitopes in a (n=14, HLA-A\*1101<sup>+</sup> patients) longitudinally studied patient cohort who received Entecavir (ETV) over the course of several years were examined. ETV is a nucleotide analogue that inhibits viral replication and leads to improved viral control in most patients. Although it does not inhibit HBeAg production by infected hepatocytes, it can also lead to HBeAg seroconversion and established anti-HBeAg antibody (HBeAb) in some patients, and this is a serological marker of further improved viral suppression. For 10 out of 14 patients who had detectable HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells (Fig. 6A), we compared the profiles of these cells in patients that later lost HBeAg and produced HBeAb (HBeAg<sup>-</sup>, n=6) over the course of the study vs. those that did not (HBeAg<sup>+</sup>, n=4). Consistent with a previous work, decreased frequencies of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells were observed in some patients (Fig. 6B and Fig. 20A). Further experiments were performed to dissect the longitudinal viral mutation and tetramer response on this given epitope (Fig. 20B), showing that these HBV-specific CD8<sup>+</sup> T cells can recognize different variants beyond the database consensus peptide sequence. Based on the analysis of these cells at multiple time-points, the detailed features could be tracked over time and major changes in the phenotypes of these cells were often observed (Fig. 6C). For instance, at early time points, HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells from patient HBeAg<sup>+</sup>03 (a patient that did not lost HBeAg during the

study time-frame) displayed a terminally differentiated effector phenotype (CD57<sup>+</sup>CD45RA<sup>+</sup>CCR7<sup>-</sup>). Whereas, at later time points, more than half of this phenotype was shift. In contrast, HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells from a HBeAg<sup>-</sup> patient (HBeAg<sup>-</sup>01, a patient that lost HBeAg and established HBeAb) displayed a memory T cell phenotype (CD27<sup>+</sup>CD127<sup>+</sup>CD45RO<sup>+</sup>) at all time points tested that was maintained for more than 6 years post treatment. Similar trends were observed for the other patients studied and detectable HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells were described quantitatively (Fig. 6D).

In addition to summarizing the composition of these cells in terms of their memory vs. effector phenotypes over time (Fig. 6D), the same markers found that vary the most across patients in our cross-sectional cohort (Fig. 1E and 2D-E) were focussed on. In general, the fraction of memory cell subset in HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells were associated with lower HBeAg level over time (Fig. 6D). This was supported by the increased expression of T cell memory-associated markers (CD27, CD127, CD45RO and CXCR3) (Fig. 6E) and the decreased expression of CD57 on HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells, in line with similar observation in InA and R patients from the cross-sectional cohort (Fig. 1E and 2D). In patients who sustained high HBeAg level, HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells displayed significantly lower levels of CD57 and higher levels of CD27 and CD127 after the virus was suppressed, whereas the cells derived from HBeAg<sup>-</sup> patients already possessed such a cellular profile (CD57<sup>lo</sup>CD27<sup>hi</sup>CD127<sup>hi</sup>CD45RO<sup>+</sup>) before the treatment and this was maintained for several years. To further quantify these changes, trajectory analysis using Scorpis was applied to these samples. To examine the reproducibility of the trajectory detection in this longitudinal cohort as it compared with the cross-sectional cohort (Fig. 2F), support vector machines (SVM) were used to map data from the longitudinal samples onto the pseudotime metric developed using Scorpis from the cross-sectional cohort (Fig. 2F, see methods). In other words, using the data from the cross-sectional cohort as a training set, we computed the pseudotime across patient's time points in the testing set (longitudinal cohort, Fig. 6F). Of these seven cellular markers, the trajectories of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells between the two independent patient cohorts were consistent (even when Scorpis was run independently without the use of SVM). In addition to validating the trajectory model on this independent cohort, this allowed the hypothesis that the HBV-specific T cell phenotypic evolution would track with the extent of viral control to be tested. Indeed, the phenotypes

of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells did show the expected progression along this pseudotime metric over the course of treatment for all patients (Fig. 6G). This implies that the virus-specific T cell response associated with viral control improved for each patient over the course of antiviral therapy, as would be expected. In addition, patients who lost HBeAg and established HBeAb had a more progressed (higher expression of T cell memory markers, lower PD-1 and TIGIT expression) phenotype at earlier time points than non-HBeAg seroconverting patients, suggesting that these patients had better virus-specific T cell response at the start of treatment. Thus, our data showed that HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells expressing increased cellular markers associated with long-term T cell memory development but decreased CD57 and two inhibitory receptor PD-1 and TIGIT was linked to viral control, and such machine learning-aided model could have predictive value for prognosis in CHB.

## Discussion

By fully leveraging a highly multiplexed combinatorial pMHC tetramer staining strategy, mass cytometry and unsupervised high-dimensional analyses, we investigated 562 unique A\*11:01-restricted candidate epitopes during the progression of HBV. Analysis of HBV-specific T cell responses is difficult due to the very low frequencies of these cells. In this regard, we show the importance of investigating both the specificity and phenotypic profiles of the antigen-specific T cells to verify their involvement in the HBV-specific immune response. Beyond this, our data highlights the heterogeneity of the virus-specific T cell response that was associated with disease stages and provides quantifiable analyses of HBV-specific TCR $\beta$  repertoires that corresponded to cellular phenotypes during chronic viral infection.

Host defense against HBV weighs on immune response driven largely by virus-specific T cells.

The number of A\*11:01-restricted epitopes detected by this comprehensive approach were relatively limited compared to the reported epitopes in the context of A\*02:01. It is possible that some epitope-specific T cells were only detectable in the liver but not in the periphery. Future investigation on HBV-specific intrahepatic lymphocytes is needed. The presence and frequencies of well-described A\*02:01-restricted HBV<sub>core18-27</sub>-specific CD8<sup>+</sup> T cells have been shown to associate with viral control. Many therapeutics have been therefore developed based on this T cells, including the blockade of overexpressed PD-1 to reinstate T cell function, adoptive transfer of engineering virus-specific T cells, and TCR-like (TCR-L) antibody to deliver interferon- $\alpha$  (IFN- $\alpha$ ) directly onto infected hepatocytes. Here, evidence was presented that showed that the specific responses and characteristics of A\*11:01-restricted HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells were linked to viral control, with public TCR clones used by these T cells. Comparative analysis showed that these cells had differing profiles across clinical stages. Furthermore, high-dimensional trajectory analysis allowed the use of the profiles of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells to assign each patient a value along an objective pseudotime metric. The underlying features of HBV<sub>core169</sub>-specific T cells along this trajectory were consistent across two independent patient cohorts, both showing correlations with viral control. Based on this metric, it is also conceivable that these antiviral-treated patients and possessing T cells with the most-progressed features of viral

control could be eligible for safe discontinuation of antiviral drug once they mounted HBV<sub>core169</sub>-specific memory T cells response, in line with a recent report showing the predictive utility of HBV-specific T cells (63). This is important because even the most advanced serological measures are unable to accurately predict such outcomes. Collectively, our findings should impact HBV immunotherapy design, and could be useful to predict patient's clinical outcome based on the phenotypic response of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells. It is also anticipated that the utility of this approach could be extended of other epitopes associated with viral control derived from HBV core or other proteins that are restricted to other HLA alleles.

By comprehensively probing HBV epitopes on numerous HBV-infected patients, the inventors here failed to identify T<sub>EX</sub> expressing all inhibitory receptors or overt evidence for "Hierarchical T cell exhaustion". Instead, unsupervised visualization using One-SENSE showed complex non-linear relationships between the expression of inhibitory receptors. Moreover, the dysfunctionality of HBV-specific T cells did not correlate to the linear accumulation of inhibitory receptors, indicating these cells were not completely functional inert. One interpretation is that these HBV-specific T cells do not nicely fit the definition of T<sub>EX</sub> as reported in LCMV-specific T cells, but instead a different type of subset that were mostly absent, with the remaining dysfunctional T cells expressing various combinations of inhibitory receptors. Based on present data, it is proposed that these T cell profiles, at least in the peripheral blood, could fit better with the description of functional adaption in CHB. Nonetheless, it is noteworthy that our functional assessment was relied on *in vitro* peptide stimulation due to the rare detection of HBV-specific T cells, and this might limit its relevance to *in vivo* response. Unlike during chronic LCMV infection, where maintenance of the T<sub>EX</sub> phenotype requires persistent and high antigen level, HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells expressing high level of CD27 and IL-7R $\alpha$  (CD127) found in InA patients were not T<sub>EX</sub> and are likely to be maintained for decades with a limited amount of viral antigen present. HBV-specific T cells from these patients were PD-1<sup>int</sup>TIGIT<sup>int</sup> with elevated expression of memory-associated markers and functional capacities that seem to be associated with control of the virus. On the other hand, for IA patients who have high and fluctuating viremia, HBV-specific T cells were detected that better matched the expected features of T<sub>EX</sub>. These included a strong co-expression of PD-1 and TIGIT and limited expression of memory-associated

markers such as CD127. Additionally, the inventors also found substantial expression of CD57 on HBV-specific T cells from IA patients, a cellular marker that is indicative of more differentiated effector cells with low proliferative capacity, suggesting these cells were not long-lived memory cells. Overall, cellular profiles of HBV-specific T cells in IA stage are consistent with persistent high antigen-exposure. At later stages (InA) in patients with better viral control, HBV-specific T cells had a memory phenotype expressing lower CD57 but elevated CXCR3, CD45RO, CD27 and CD127. The lack of the detectable HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells in IT patients who has high viremia suggests that they might be largely deleted, and the absence of such particular T cells could contribute to the minimum liver inflammation in this stage. Future studies involve larger volume of blood samples from IT patients or using more sensitive approaches may help to address this issue. Further investigation of the expression level of EOMES and T-bet, or the epigenetic modification that better defined the bona fide T<sub>EX</sub> is important to address this aspect in CHB. It is also important to note that this report is limited to the analysis of circulating HBV-specific CD8<sup>+</sup> T cells, further examination of the exhaustion profile in intrahepatic lymphocytes is needed to address this question.

Previous studies have suggested the link between the different inhibitory receptors and T cell differentiation, which is in relative disagreement with the present One-SENSE analysis objectively showing more complex relationships between the co-expressed inhibitory receptors and T cell differentiation on several virus-specific T cells across multiple clinical stages of HBV infection. This is because many studies in human chronic viral infection often examined less than four inhibitory receptors within T cell subsets that were simply defined by few differentiation-associated markers, or on limited numbers of virus-specific T cells from one type of patient. Secondly, traditional analysis using hierarchical gating on biaxial dot plots to assess the expression level of cellular protein can easily underestimate the phenotypic complexity.

Despite the TCR sequence diversity of T cells specific for the HBV<sub>core169</sub> epitope, several public clones at relatively high abundance were detected in multiple patients. In line with the CMV-specific TCR repertoire, it was noted that these previously unidentified public TCRs were different when derived from patients showing viral control (R and InA) vs. viremia (IA),

suggesting the functional importance of these T cells. The public virus-specific TCR clones may be selected over the course of viral clearance (i.e. from IA into InA) and contraction of the effector response. As previously reported for CMV and EBV infection, virus-specific CD8<sup>+</sup> T cells do not express IL-7R $\alpha$  (CD127) until T cell memory had been established, and such selection is thought to be driven by high affinity TCR-viral epitope binding. This is consistent with the characteristics of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells in InA and R patients who carried public TCRs and elevated expression of T cell memory-associated markers including CD127, indicating their long-lived and self-renewing ability to maintain memory T cells pool after the reduction of viral antigen. Analogously, intrahepatic and peripheral public TCR clones have been linked to viral clearance in HCV-infected chimpanzees.

Despite the challenges highlighted associated with detecting HBV-specific T cells due to their low prevalence, the present invention explores the previously unappreciated complexity of virus-specific T cells in lifelong human HBV viral infection. The cellular responses of HBV<sub>core169</sub>-specific CD8<sup>+</sup> T cells and TCR sequences used were associated with the status of HBV infection and could be used as an indicator of the relative extent of viral control. Thus, the results provided here could have important implications for the development of new biomarkers, treatment strategies and immunotherapy aiming at HBV cure.

Whilst there has been described in the foregoing description preferred embodiments of the present invention, it will be understood by those skilled in the technology concerned that many variations or modifications in details of design or construction may be made without departing from the present invention.

## CLAIMS

1. Use of a peptide for identifying Hepatitis B virus antigen-specific T cells, the peptide is  
5 capable of binding HLA-A\*1101 and when bound to HLA-A\*1101 is capable of identifying T cells specific for Hepatitis B virus, wherein the peptide comprises an amino acid sequence selected from the group consisting of STLPETAVVRR, STLPETAVVR, STLPETTVIRR, STPPETTVVRR, STLPETTVVGR and STIPETTVVRR.
- 10 2. A method for selecting Hepatitis B virus antigen-specific T cells, the method comprising contacting a population of T cells with a peptide comprising an amino acid sequence selected from the group consisting of STLPETAVVRR, STLPETAVVR, STLPETTVIRR, STPPETTVVRR, STLPETTVVGR and STIPETTVVRR.
- 15 3. An expression vector comprising a polynucleotide encoding a T cell receptor (TCR) molecule, the TCR molecule comprising an amino acid sequence selected from the group consisting of: CASGDSNSPLHF, CASSGGQIVYEQYF, CSARGGRGGDYTF and CASSQDWTEAFF.
4. The expression vector of claim 3, wherein the polynucleotide encoding the TCR  
20 molecule has a sequence selected from the group consisting of SEQ ID NOs 1 to 16.
5. A host cell comprising:
- (i) a polynucleotide encoding a TCR molecule, the TCR molecule comprising an amino acid sequence selected from the group comprising: CASGDSNSPLHF, CASSGGQIVYEQYF,  
25 CSARGGRGGDYTF and CASSQDWTEAFF; or
- (ii) a polynucleotide encoding a TCR molecule comprising a sequence selected from the group consisting of SEQ ID NOs 1 to 16; or
- (iii) an expression vector according to claim 3; or
- (iv) an expression vector according to claim 4.
- 30 6. The host cell according to claim 5, wherein the host cell is a T cell derived from a patient.

7. A T cell modified to express the TCR molecule comprising an amino acid sequence selected from the group consisting of: CASGDSNSPLHF, CASSGGQIVYEQYF, CSARGGRGGDYTF and CASSQDWTEAFF.

5

8. The T cell according to claim 7, wherein the amino acid sequence encodes for a TCR beta chain portion.

9. The T cell according to claims 7 or 8, for use in inducing antiviral T cell activity against a Hepatitis B virus core169 epitope, wherein a polynucleotide or an expression vector according to claims 3 or 4 has been introduced into the T cell, preferably patient-derived T cell, so that the T cell expresses the encoded TCR molecule, wherein the polynucleotide encodes a TCR molecule, wherein the TCR molecule comprises an amino acid sequence selected from the group comprising: CASGDSNSPLHF, CASSGGQIVYEQYF, CSARGGRGGDYTF and CASSQDWTEAFF, and wherein the polynucleotide has a sequence selected from the group comprising SEQ ID NOs 1 to 16.

10. A pharmaceutical composition comprising:  
(i) a peptide comprising an amino acid sequence selected from the group consisting of STLPETAVVRR, STLPETAVVR, STLPETTVIRR, STPPETTVVRR, STLPETTVVGR and STIPETTVVRR; or  
(ii) a polynucleotide encoding a TCR molecule, the TCR molecule comprising an amino acid sequence selected from the group comprising: CASGDSNSPLHF, CASSGGQIVYEQYF, CSARGGRGGDYTF and CASSQDWTEAFF; or  
(iii) a T cell according to claim 7;  
and a pharmaceutically acceptable carrier.

11. A peptide comprising an amino acid sequence selected from the group consisting of STLPETAVVRR, STLPETAVVR, STLPETTVIRR, STPPETTVVRR, STLPETTVVGR and STIPETTVVRR; or a T cell according to claim 7;  
for use in medicine, wherein an individual to be treated is a human individual who carries HLA-A\*1101.

12. A vaccine against Hepatitis B virus infection comprising a peptide comprising an amino acid sequence selected from the group consisting of STLPETAVVRR, STLPETAVVR, STLPETTVIRR, STPPETTVVRR, STLPETTVVGR and STIPETTVVRR.

5 13. Use of:

a peptide comprising an amino acid sequence selected from the group consisting of STLPETAVVRR, STLPETAVVR, STLPETTVIRR, STPPETTVVRR, STLPETTVVGR and STIPETTVVRR; or  
a T cell according to claim 7;

in the manufacture of a medicament for treating a Hepatitis B virus infection.

10

14. Use of a T cell according to claim 7 in the preparation of a medicament for inducing antiviral T cell activity against a Hepatitis B virus.

15