

CLAIMS

1. A method of identifying one or more shared candidate antigens for characterising and/or treating a medical condition, the method including:
 - (i) obtaining transcriptomic data for test samples from a first cohort of patients having the medical condition;
 - (ii) obtaining reference transcriptomic data for a set of reference samples;
 - (iii) determining, by a comparison of the transcriptomic data to the reference transcriptomic data, one or more splice variants that are more highly transcribed in each sample of a subset of the test samples as compared to the reference samples;
 - (iv) determining, for each said shared splice variant, one or more amino acid sequences that occur in an amino acid translation of the shared splice variant, but not in amino acid translations of corresponding splice variants of the same gene that are transcribed in the reference samples; and
 - (v) predicting HLA binding of the one or more shared amino acid sequences, or part thereof, to identify the one or more shared amino acid sequences as one or more shared candidate antigens.
2. A method according to claim 1, further including determining for each said shared splice variant, prior to step (iv), whether there is a change in reading frame in the first shared splice variant relative to the one or more corresponding splice variants of the same gene.
3. A method according to claim 1 or claim 2, wherein step (iv) includes determining non-overlapping nucleotide sequence between the shared splice variant and corresponding splice variants of the same gene.
4. A method according to any one of the preceding claims, wherein the subset comprising the shared splice variant comprises more than a threshold number or more than a threshold percentage of the test samples.
5. A method according to any one of the preceding claims, further including determining

one or more common HLA alleles that occur in more than a predetermined proportion of the first cohort of patients; wherein step (v) includes: generating a plurality of candidate peptides from the one or more amino acid sequences; and predicting binding of the plurality of shared candidate peptides to proteins encoded by the one or more common HLA alleles.

6. A method according to any one of the preceding claims, wherein the set of reference samples includes matched normal samples from the first cohort of patients.
7. A method according to any one of the preceding claims, wherein the set of reference samples includes samples from a cohort of subjects who do not have the medical condition.
8. A method according to any one of the preceding claims, wherein the medical condition is cancer.
9. A method according to claim 8, wherein the medical condition is gastric cancer, head and neck cancer, colorectal cancer or hepatocellular cancer.
10. A method according to any one of the preceding claims, wherein the first cohort of patients is of a specific HLA subtype.
11. A method according to any one of the preceding claims, wherein the medical condition is common to a group of patients.
12. A method according to any one of the preceding claims, wherein the shared candidate antigen is a MARK3, NBPF9, PARD3, ZC3HAV1, YAF2, CAMKK1, LRR1, ZNG670, GRINA or MZF1 splice variant and the HLA allele is HLA-A11 or HLA-A24.
13. A method according to claim 12, wherein the splice variant comprises a peptide having at least 80% sequence identity to any one of SEQ ID NO: 1, 31-38, 51 or 52, or a nucleic acid encoding a peptide having at least 80% sequence identity to any one of SEQ ID NO: 1, 31-38, 51 or 52.

14. A method according to any one of the preceding claims, wherein the method comprises verifying or testing HLA binding of the one or more shared amino acid sequences to identify the one or more shared amino acid sequences as a shared candidate antigens.
15. A method of identifying a shared antigen-T lymphocyte pair, the method comprising:
 - a) identifying a shared candidate antigen according to any one of claims 1 to 14; providing one or more respective labelled biomolecules comprising a label and a peptide comprising the shared candidate antigen;
 - b) contacting the one or more labelled biomolecules with one or more samples containing peripheral blood from patients having the medical condition; and
 - c) identifying, from the one or more samples, T lymphocytes that are bound to said labelled biomolecules, so as to identify a shared antigen-T lymphocyte pair.
16. A method according to claim 15, wherein the labelled biomolecules comprise HLA multimers.
17. A method according to claim 15 or claim 16, wherein labelled biomolecules containing respective shared candidate antigens are labelled with different respective barcodes.
18. A method according to claim 17, wherein the barcodes are heavy metal barcodes.
19. A method according to any one of claims 15 to 18, wherein the respective patients having the medical condition are part of a second cohort of patients that does not overlap with the first cohort of patients.
20. A method according to any one of claims 15 to 19, wherein the method comprises testing the biological function of the T lymphocytes.
21. A method of claim 20, wherein the method comprises characterising the T lymphocytes to determine whether they are cytotoxic and/or testing whether the shared antigens are immunogenic.
22. A method for identifying T lymphocytes that bind specifically to one or more shared

candidate antigens identified according to any one of claims 1 to 14, comprising:

- (i) providing one or more respective labelled biomolecules comprising a label and a respective shared candidate antigen;
- (ii) contacting the one or more labelled biomolecules with one or more samples containing peripheral blood from respective patients having the medical condition; and
- (iii) identifying, from the one or more samples, T lymphocytes that are bound to said labelled biomolecules.

23. A method according to claim 22, wherein identification of T lymphocytes that are bound to said labelled biomolecules characterises the respective patients as having a medical condition that is associated with the expression of the one or more shared antigens.

24. A method according to claim 22 or claim 23, wherein the method comprises testing the biological function of the T lymphocytes.

25. A method of claim 24, wherein the method comprises characterising the T lymphocytes to determine whether they are cytotoxic and/or testing whether the shared antigens are immunogenic.

26. A method of characterising a medical condition in a subject, the method comprising determining the level of one or more shared antigens identified according to any one of claims 15 to 21, wherein an increased level of the one or more shared antigens as compared to a reference characterises the medical condition as one that is associated with the expression of the one or more shared antigens.

27. A method of treating a medical condition in a subject, the method comprising:

- (a) determining the level of one or more shared antigens identified according to any one of claims 15 to 21, wherein an increased level of the one or more shared antigens as compared to a reference characterises the medical condition in the subject as one that is associated with the expression of one or more shared antigens, and
- (b) treating the subject found to have a medical condition that is associated with the expression of the one or more shared antigens.

28. A method of characterising a medical condition in a subject, the method comprising determining the level of T lymphocytes that bind specifically to one or more shared antigens identified according to any one of claims 15 to 21, wherein an increased level of the T lymphocytes as compared to a reference characterises the medical condition as one that is associated with the expression of the one or more shared antigens.
29. A method of treating a medical condition in a subject, the method comprising a) determining the level of T lymphocytes that binds specifically to one or more shared antigens identified according to any one of claims 15 to 21, wherein an increased level of the T lymphocytes as compared to a reference characterises the medical condition as one that is associated with the expression of the one or more shared antigens; and b) treating the subject found to have a medical condition that is associated with the expression of the one or more shared antigens.
30. A method of treating a medical condition in a subject, the method comprising:
- (a) determining the level of T lymphocytes that binds specifically to one or more shared antigens identified according to any one of claims 15 to 21, wherein an increased level of the T lymphocytes as compared to a reference characterises the medical condition as one that expresses the one or more shared antigens;
 - (b) isolating and expanding the population of T lymphocytes *ex vivo*; and
 - (c) administering the expanded population of T lymphocytes to the subject to treat the medical condition in the subject found to express the one or more shared antigens.
31. A method of treating a medical condition in a subject, the method comprising:
- (a) isolating a population of T lymphocytes that binds specifically to one or more shared antigens identified according to any one of claims 15 to 21 in a subject suffering from the medical condition, and expanding the population of T lymphocytes *ex vivo*; and
 - (b) administering the expanded population of T lymphocytes to the subject to treat the medical condition in the subject.
32. An immunomodulatory composition comprising one or more shared antigen identified according to any one of claims 15 to 21 and a pharmaceutically acceptable carrier.

33. A composition according to claim 32, wherein the shared antigen is a peptide having at least 80% sequence identity to any one of SEQ ID NOs: 1, 31-38, 51 or 52, or is a nucleic acid encoding a peptide having at least 80% sequence identity to any one of SEQ ID NOs: 1 31-38, 51 or 52.
34. A method of stimulating an immune response in a subject, the method comprising administering a effective amount of an immunomodulatory composition according to claim 32 or claim 33 to the subject under conditions and for a sufficient time to stimulate the immune response in the subject.
35. A shared antigen-T lymphocyte pair identified according to any one of claims 15 to 21, wherein the shared antigen is a MARK3, NBPF9, PARD3, ZC3HAV1, YAF2, CAMKK1, LRR1, ZNG670, GRINA or MZF1 splice variant, the HLA subtype is HLA-A11 or HLA-A24, and the T lymphocyte binds to the shared antigen.
36. A labelled biomolecule comprising a HLA molecule bound to a shared antigen identified according to any one of claims 15 to 21 for use in detecting the presence or determining the level of T lymphocytes that binds specifically to the shared antigen.
37. A labelled biomolecule according to claim 36, wherein the shared antigen is a peptide having at least 80% sequence identity to any one of SEQ ID NO:1, 31-38, 51 or 52, or is a nucleic acid encoding a peptide having at least 80% sequence identity to any one of SEQ ID NO:1, 31-38, 51 or 52, and wherein the HLA is HLA-A11 or HLA-A24.
38. An antibody that binds specifically to a shared antigen identified according to any one of claims 15 to 21, wherein the shared antigen is bound to a HLA molecule.
39. An antibody of claim 38, wherein the shared antigen is a peptide having at least 80% sequence identity to any one of SEQ ID NO:1, 31-38, 51 or 52, or is a nucleic acid encoding a peptide having at least 80% sequence identity to any one of SEQ ID NO:1, 31-38, 51 or 52.
40. A T-cell receptor (TCR) that binds to a shared antigen according to any one of claims 15 to 21, wherein the shared antigen is bound to a HLA molecule.

41. A TCR of claim 40, wherein the shared antigen is a peptide having at least 80% sequence identity to any one of SEQ ID NO: 1, 31-38, 51 or 52, or a nucleic acid encoding a peptide having at least 80% sequence identity to any one of SEQ ID NO: 1, 31-38, 51 or 52.
42. The TCR of claim 40 or claim 41, wherein the TCR comprises a) a TCR α chain variable domain comprising a sequence having at least 70% sequence identity to SEQ ID NO: 21, and b) a TCR β chain variable domain comprising a sequence having at least 70% sequence identity to SEQ ID NO: 29.
43. The TCR of any one of claims 40 to 42, wherein TCR comprises a) a TCR α chain domain comprising i) a CDR1 sequence of SEQ ID NO: 17, ii) a CDR2 sequence of SEQ ID NO: 18 and iii) a CDR3 of SEQ ID NO: 20; and b) a TCR β chain domain comprising i) a CDR1 sequence of SEQ ID NO: 25, ii) a CDR2 sequence of SEQ ID NO: 26 and/or iii) a CDR3 sequence of SEQ ID NO: 28.
44. The TCR of any one of claims 40 to 43, wherein the TCR is a solubilised TCR.
45. An engineered immune cell comprising a nucleic acid encoding a T-cell receptor of any one of claims 40 to 43, wherein the engineered immune cell is capable of specifically binding to a shared antigen or fragment thereof, wherein the shared antigen or fragment thereof is bound to a HLA molecule.
46. A method of treating a medical condition in a subject, the method comprising administering a solubilised TCR of claim 44 or an engineered immune cell of claim 45 to the subject for a sufficient time and under conditions to treat the medical condition in the subject.
47. A method of producing an antibody, the method comprising:
- (a) immunizing an animal with a shared antigen identified according to any one of claims 15 to 21;
 - (b) identifying and/or isolating a B cell from the animal, which binds specifically to the antigen ; and
 - (c) producing the antibody expressed by that B cell.

48. A pharmaceutical composition comprising an antibody of claim 38 or claim 39 or a soluble TCR of claim 44, or an engineered immune cell of claim 45, and a pharmaceutically acceptable carrier.
49. A method of treating a medical condition in a subject, the method comprising administering a pharmaceutical composition of claim 48 to the subject for a sufficient time and under conditions to treat the medical condition in the subject.
50. A method of identifying a shared antigen-T lymphocyte pair, the method comprising:
- (i) obtaining transcriptomic data for test samples from a first cohort of patients having the medical condition, wherein the cohort comprises a plurality of patients;
 - (ii) obtaining reference transcriptomic data for a set of reference samples;
 - (iii) determining, by a comparison of the transcriptomic data to the reference transcriptomic data, one or more splice variants that are more highly transcribed in each sample of a subset of the test samples as compared to the reference samples,
 - (iv) determining, for each said shared splice variant, one or more amino acid sequences that occur in an amino acid translation of the splice variant, but not in amino acid translations of corresponding splice variants of the same gene that are transcribed in the reference samples;
 - (v) predicting HLA binding of the one or more shared amino acid sequences, or part thereof, to identify the one or more amino acid sequences as one or more shared candidate antigens;
 - (vi) providing one or more labelled biomolecules comprising a label and a peptide comprising a shared candidate antigen;
 - (vii) contacting the one or more labelled biomolecules with one or more samples containing peripheral blood from patients having the medical condition; and
 - (viii) identifying, from the one or more samples, T lymphocytes that are bound to said labelled biomolecules, so as to identify a shared antigen-T lymphocyte pair.