
Antigen Reactive T-Cell Receptors

5 The present invention relates to a method of identifying a T-cell reactive to cells of a subject presenting a T-cell activating antigen (reactive T-cell), comprising (a) determining expression of at least one of CCL4, CCL4L2, CCL3, CCL3L1, and CXCL13 in T-cells from a sample of said subject; and (b) identifying a reactive T-cell based on the determination of step (a). The present invention also relates to a method of identifying a TCR binding to an activating antigen
10 presented on a cell, preferably a cancer cell, of a subject, said method comprising (A) identifying a reactive T-cell according to the method of identifying a reactive T-cell, (B) providing the amino acid sequences of at least the complementarity determining regions (CDRs) of the TCR of the reactive T-cell identified in step (A); and, hereby, (C) identifying a TCR binding to an activating antigen presented on a cell. The present invention further relates to further methods
15 and cancer-reactive T-cells related thereto.

Over recent years, there has been increasing interest in identifying antigen-reactive T-cell receptors (TCRs) for personalized Adoptive Cell Therapies (ACT). In such a therapy, a patient's circulating T cells in the blood are harvested, transgenically modified to express a tumor reactive TCR, and then reinfused into the patient.
20

As a source of T-cells for identifying e.g. tumor-reactive TCRs, tumor-infiltrating lymphocytes (TILs) have been used. Tumor reactive T cells within a TIL population can in theory be identified by their upregulation of known T cell activation biomarkers such as CD69 and Nur77, though in practice the value of TCRs identified by such an approach has been limited.
25

Further biomarkers of T-cell activation have been described, cf. Cano-Gamez et al. (2020), Nat Comm 11:, art. 1801 (doi.org/10.1038/s41467-020-15543-y), Magen et al. (2019), Cell Rep 29(10):3019 (doi.org/10.1016/j.celrep.2019.10.131), and Oh et al. (2020), Cell 181(7):1612 (doi.org/10.1016/j.cell.2020.05.017). Moreover, e.g. biomarkers predicting non-response to
30 immune checkpoint blockade (WO2018/209324) and biomarkers for immunotherapy resistance

(WO2019/070755) have been described. Recently, activation markers from tumor infiltrating T lymphocytes were described (WO 2021/188954 A1, Lowery et al. (2022), Science 10.1126/science.abl5447).

5 T-cell activation has long been acknowledged to involve presentation of an antigen, e.g. an epitope of a polypeptide, in the context of major histocompatibility complexes (MHCs). MHC class I, interacting with TCR complexes comprising the CD8 protein on CD8⁺ T-cells, is expressed by all nucleate cells, while MHC class II, interacting with TCR complexes comprising the CD4 protein on CD4⁺ T-cells, is only expressed by professional antigen presenting cells,
10 mostly B- cells and dendritic cells. However, other surface molecules of cells have been found to be involved in T-cell interaction and activation as well (cf. e.g. Iwabuchi & van Kaer (2019), Front Immunol 10:1837 (doi: 10.3389/fimmu.2019.01837)).

Nonetheless, there is still a need for improved methods for providing T-cells reactive to specific
15 antigens, e.g. cancer antigens, and corresponding TCRs. This problem is solved by the embodiments characterized in the claims and described herein below.

In accordance, the present invention relates to a method of identifying a T-cell reactive to cells of a subject presenting a T-cell activating antigen (reactive T-cell), comprising
20 (a) determining expression of at least one of CCL4, CCL4L2, CCL3, CCL3L1, and CXCL13 in T-cells from a sample of said subject; and
(b) identifying a reactive T-cell based on the determination of step (a).

Preferably, the present invention relates to a method of identifying a T-cell reactive to cancer
25 cells of a subject (cancer-reactive T-cell), comprising
(a) determining expression of at least one of CCL4, CCL4L2, CCL3, CCL3L1, and CXCL13 in T-cells from a sample of said subject; and
(b) identifying a cancer-reactive T-cell based on the determination of step (a).

30 In general, terms used herein are to be given their ordinary and customary meaning to a person of ordinary skill in the art and, unless indicated otherwise, are not to be limited to a special or customized meaning. As used in the following, the terms “have”, “comprise” or “include” or any arbitrary grammatical variations thereof are used in a non-exclusive way. Thus, these terms

may both refer to a situation in which, besides the feature introduced by these terms, no further features are present in the entity described in this context and to a situation in which one or more further features are present. As an example, the expressions "A has B", "A comprises B" and "A includes B" may both refer to a situation in which, besides B, no other element is present in A (i.e. a situation in which A solely and exclusively consists of B) and to a situation in which, besides B, one or more further elements are present in entity A, such as element C, elements C and D or even further elements. Also, as is understood by the skilled person, the expressions "comprising a" and "comprising an" preferably refer to "comprising one or more", i.e. are equivalent to "comprising at least one". In accordance, expressions relating to one item of a plurality, unless otherwise indicated, preferably relate to at least one such item, more preferably a plurality thereof: thus, e.g. identifying "a cell" relates to identifying at least one cell, preferably to identifying a multitude of cells.

Further, as used in the following, the terms "preferably", "more preferably", "most preferably", "particularly", "more particularly", "specifically", "more specifically", or similar terms are used in conjunction with optional features, without restricting further possibilities. Thus, features introduced by these terms are optional features and are not intended to restrict the scope of the claims in any way. The invention may, as the skilled person will recognize, be performed by using alternative features. Similarly, features introduced by "in an embodiment", "in a further embodiment", or similar expressions are intended to be optional features, without any restriction regarding further embodiments of the invention, without any restrictions regarding the scope of the invention and without any restriction regarding the possibility of combining the features introduced in such way with other optional or non-optional features of the invention.

As used herein, the term "standard conditions", if not otherwise noted, relates to IUPAC standard ambient temperature and pressure (SATP) conditions, i.e. preferably, a temperature of 25°C and an absolute pressure of 100 kPa; also preferably, standard conditions include a pH of 7. Moreover, if not otherwise indicated, the term "about" relates to the indicated value with the commonly accepted technical precision in the relevant field, preferably relates to the indicated value $\pm 20\%$, more preferably $\pm 10\%$, most preferably $\pm 5\%$. Further, the term "essentially" indicates that deviations having influence on the indicated result or use are absent, i.e. potential deviations do not cause the indicated result to deviate by more than $\pm 20\%$, more preferably $\pm 10\%$, most preferably $\pm 5\%$. Thus, "consisting essentially of" means including the components

- specified but excluding other components except for materials present as impurities, unavoidable materials present as a result of processes used to provide the components, and components added for a purpose other than achieving the technical effect of the invention. For example, a composition defined using the phrase "consisting essentially of" encompasses any known acceptable additive, excipient, diluent, carrier, and the like. Preferably, a composition consisting essentially of a set of components will comprise less than 5% by weight, more preferably less than 3% by weight, even more preferably less than 1% by weight, most preferably less than 0.1% by weight of non-specified component(s).
- 10 The degree of identity (e.g. expressed as "%identity") between two biological sequences, preferably DNA, RNA, or amino acid sequences, can be determined by algorithms well known in the art. Preferably, the degree of identity is determined by comparing two optimally aligned sequences over a comparison window, where the fragment of sequence in the comparison window may comprise additions or deletions (e.g., gaps or overhangs) as compared to the sequence
- 15 it is compared to for optimal alignment. The percentage is calculated by determining, preferably over the whole length of the polynucleotide or polypeptide, the number of positions at which the identical residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Optimal
- 20 alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman (1981), by the homology alignment algorithm of Needleman and Wunsch (1970), by the search for similarity method of Pearson and Lipman (1988), by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, PASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr.,
- 25 Madison, WI), or by visual inspection. Given that two sequences have been identified for comparison, GAP and BESTFIT are preferably employed to determine their optimal alignment and, thus, the degree of identity. Preferably, the default values of 5.00 for gap weight and 0.30 for gap weight length are used. In the context of biological sequences referred to herein, the term "essentially identical" indicates a %identity value of at least 80%, preferably at least 90%, more
- 30 preferably at least 98%, most preferably at least 99%. As will be understood, the term essentially identical includes 100% identity. The aforesaid applies to the term "essentially complementary" mutatis mutandis.

The term "fragment" of a biological macromolecule, preferably of a polynucleotide or polypeptide, is used herein in a wide sense relating to any sub-part, preferably subdomain, of the respective biological macromolecule comprising the indicated sequence, structure and/or function. Thus, the term includes sub-parts generated by actual fragmentation of a biological macromolecule, but also sub-parts derived from the respective biological macromolecule in an abstract manner, e.g. in silico. In the context of sequence information, in particular nucleic acid sequences and/or polypeptide sequences, the term "sub-sequence" is used for sequences representing only a part of a longer sequence .

10 Unless specifically indicated otherwise herein, the compounds specified, in particular polynucleotides, polypeptides, or fragments thereof, e.g. variable regions of a T-cell receptor (TCR), may be comprised in larger structures, e.g. may be covalently or non-covalently linked to accessory molecules, carrier molecules, retardants, and other excipients. In particular, polypeptides as specified may be comprised in fusion polypeptides comprising further peptides, which
15 may serve e.g. as a tag for purification and/or detection, as a linker, or to extend the in vivo half-life of a compound. The term "detectable tag" refers to a stretch of amino acids which are added to or introduced into the fusion polypeptide; preferably, the tag is added C- or N- terminally to the fusion polypeptide of the present invention. Said stretch of amino acids preferably allows for detection of the fusion polypeptide by an antibody which specifically recognizes the
20 tag; or it preferably allows for forming a functional conformation, such as a chelator; or it preferably allows for visualization, e.g. in the case of fluorescent tags. Preferred detectable tags are the Myc-tag, FLAG-tag, 6-His-tag, HA-tag, GST-tag or a fluorescent protein tag, e.g. a GFP-tag. These tags are all well known in the art. Other further peptides preferably comprised in a fusion polypeptide comprise further amino acids or other modifications which may serve as
25 mediators of secretion, as mediators of blood-brain-barrier passage, as cell-penetrating peptides, and/or as immune stimulants. Further polypeptides or peptides to which the polypeptides may be fused are signal and/or transport sequences and/or linker sequences. A variable region of a TCR, preferably, is comprised in a backbone of a TCR alpha or beta chain as specified herein below.

30

The term "polypeptide", as used herein, refers to a molecule consisting of several, typically at least 20 amino acids that are covalently linked to each other by peptide bonds. Molecules consisting of less than 20 amino acids covalently linked by peptide bonds are usually considered

to be "peptides". Preferably, the polypeptide comprises of from 50 to 1000, more preferably of from 75 to 750, still more preferably of from 100 to 500, most preferably of from 110 to 400 amino acids. Preferably, the polypeptide is comprised in a fusion polypeptide and/or a polypeptide complex.

5

The method of identifying a reactive T-cell of the present invention, preferably, is an in vitro method. The method may comprise further steps in addition to those related to herein above. For example, further steps may relate, e.g., to providing a sample for step a), or determining further biomarkers in step b). Moreover, one or more of said steps may be performed or assisted by automated equipment.

10

The term "T-cell receptor", abbreviated as "TCR", as used herein, relates to a polypeptide complex on the surface of T-cells mediating recognition of antigenic peptides presented by target cells, preferably in the context of MHC molecules or MHC-related molecules such as MR1 or CD1, more preferably in the context of MHC molecules, still more preferably in the context of MHC class I or MHC class II molecules, most preferably in the context of MHC class I molecules. Typically, the TCR comprises one TCR-alpha chain and one TCR-beta chain, i.e. is an alpha/beta chain heterodimer. The TCR may, however, also comprise a TCR gamma and a TCR delta chain instead of the TCR alpha and beta chains. The TCR alpha and beta or gamma and delta chains mediate antigen recognition and each comprise a transmembrane region, a constant region, a joining region, and a variable region, the variable region of each TCR alpha, beta, gamma, or delta chain comprising three complementarity determining regions (CDRs), referred to as CDR1, CDR2, and CDR3, respectively. In accordance with usual nomenclature, the complex consisting of an alpha and a beta chain or a gamma and a delta chain is referred to as "T-cell receptor" or "TCR" herein, the alpha and/or beta chain and the gamma and/or delta chains commonly or singly being referred to a "TCR polypeptide" or "TCR polypeptides", whereas the polypeptide complex comprising a TCR and accessory polypeptides, such as CD3 and CD247, is referred to as "T-cell receptor complex", abbreviated as "TCR complex". Preferably the T-cell receptor binds to a major histocompatibility complex (MHC) molecule, preferably an MHC class I or class II, more preferably an MHC class I molecule, presenting an antigen contributing and/or associated with disease, preferably a cancer antigen or an autoimmune T-cell antigen, more preferably a cancer antigen, still more preferably an epitope of a cancer specific antigen, in particular a neoepitope of a cancer cell. Binding of a T-cell receptor to an

15

20

25

30

antigen can be determined by methods known to the skilled person, e.g. by methods as specified herein in the Examples, or e.g. in a tetramer assay. Preferably, binding of the TCR to an epitope presented on an MHC activates the T-cell. Activation biomarkers of various types of T-cells are known in the art and include in particular CD69, CD137, CD27, TRAP/CD40L, and CD134.

5 The TCR may also be a soluble TCR. The term "soluble TCR" is, in principle, known to the skilled person to relate to a TCR as specified herein above lacking the transmembrane domains. Thus, preferably, the soluble TCR comprises the constant and the variable regions of the TCR polypeptides of a TCR. More preferably, the soluble TCR comprises the variable regions of the TCR polypeptides of a TCR, preferably in the form of a fusion polypeptide.

10

The term "complementarity determining region", abbreviated as "CDR", is understood by the skilled person. As is known in the art, each TCR alpha, beta, gamma, and delta chain comprises three CDRs, which are the peptides providing the epitope-specificity determining contacts of a TCR to a peptide presented by an MHC molecule as specified elsewhere herein.

15

The term "T-cell" is understood by the skilled person to relate to a lymphocyte expressing at least one type of T-cell receptor as specified herein above. Preferably, the T-cell is a CD8⁺ T-cell recognizing MHC class I molecules on the surface of target cells, or is a CD4⁺ T-cell recognizing MHC class II molecules on the surface of target cells, more preferably is a CD8⁺

20

T-cell. Preferably, the T-cell is a cytotoxic T-cell, more preferably a CD8⁺ cytotoxic T-cell, which may also be referred to as "killer cell". Also preferably, the T-cell is a regulatory or helper T-cell, more preferably a regulatory T-cell. Preferably, the T-cell is an alpha/beta T-cell, i.e. a T-cell expressing a T-cell receptor comprising a TCR alpha and a TCR beta chain. Preferably, the T-cell is reactive to cells presenting a T-cell activating antigen, i.e. is a "reactive T-

25

cell", more preferably is specifically reactive to cells presenting a T-cell activating antigen; thus, the T-cell preferably is activated by cells presenting a T-cell activating antigen, preferably is specifically activated by cells presenting a T-cell activating antigen, the terms "specifically activated by cells presenting a T-cell activating antigen" and "specifically reactive to cells presenting a T-cell activating antigen" indicating that the T-cell preferably is activated by cells presenting a T-cell activating antigen, but not by cells not presenting a T-cell activating antigen,

30

in particular of the same tissue. Activation of T-cells can be measured by methods known in the art, e.g. by measuring cytokine secretion, e.g. interferon-gamma secretion, or by a method as specified herein in the Examples. Preferably, the T-cell is reactive to cancer cells, i.e. is a

"cancer-reactive T-cell" or is reactive to cells presenting a T-cell autoantigen, i.e. is an "auto-immune-reactive T-cell". Thus, preferably, the T-cell expresses a TCR recognizing a cancer antigen, preferably a cancer-specific antigen, as specified herein below. In accordance with the above, a T-cell reactive to cancer cells is a T-cell expressing a TCR recognizing a cancer anti-
5 gen, preferably a cancer-specific antigen. Also preferably, the T-cell expresses a TCR recognizing an autoimmune T-cell antigen, preferably a specific autoimmune T-cell antigen.

The term "T-cell activating antigen", for which also the expression "activating antigen" may be used, is used herein in a broad sense to relate to any structure presented on the surface of a cell
10 of a subject which can activate a T-cell expressing an appropriate TCR. Preferably, the antigen is a polypeptide or fragment thereof, a polysaccharide, or a lipid. More preferably, the antigen is an epitope of a polypeptide presented by said cell of said subject in the context of an MHC molecule, preferably as specified herein above. As the skilled person understands, if a reactive T-cell is identified by the method as specified herein in a sample, there preferably is a presump-
15 tion that there are cells in said subject presenting a T-cell activating antigen; since this identification not necessarily includes identifying the T-cell activating antigen, the reactive T-cell identified and/or its TCR may be used further for identifying the T-cell activating antigen. Preferably, the T-cell activating antigen is a cancer antigen or an autoimmune-related T-cell activating antigen. Thus, the reactive T-cell may in particular be a cancer-reactive T-cell or an autoim-
20 mune-reactive T-cell.

The term "cancer", as used herein, relates to a disease of an animal, including man, characterized by uncontrolled growth by a group of body cells ("cancer cells"). This uncontrolled growth may be accompanied by intrusion into and destruction of surrounding tissue and possibly spread
25 of cancer cells to other locations in the body. Preferably, also included by the term cancer is a relapse. Thus, preferably, the cancer is a solid cancer, a metastasis, or a relapse thereof. Cancer may be induced by an infectious agent, preferably a virus, more preferably an oncogenic virus, more preferably Epstein-Barr virus, a hepatitis virus, Human T-lymphotropic virus 1, a papillomavirus, or Human herpesvirus 8. Cancer may, however, also be induced by chemical com-
30 pounds, e.g. a carcinogen, or endogenously, e.g. caused by spontaneous mutation.

Preferably, the cancer is selected from the list consisting of acute lymphoblastic leukemia, acute myeloid leukemia, adrenocortical carcinoma, aids-related lymphoma, anal cancer, appendix

cancer, astrocytoma, atypical teratoid, basal cell carcinoma, bile duct cancer, bladder cancer, brain stem glioma, breast cancer, burkitt lymphoma, carcinoid tumor, cerebellar astrocytoma, cervical cancer, chordoma, chronic lymphocytic leukemia, chronic myelogenous leukemia, colon cancer, colorectal cancer, craniopharyngioma, endometrial cancer, ependymblastoma, ependymoma, esophageal cancer, extracranial germ cell tumor, extragonadal germ cell tumor, extrahepatic bile duct cancer, fibrosarcoma, gallbladder cancer, gastric cancer, gastrointestinal stromal tumor, gestational trophoblastic tumor, hairy cell leukemia, head and neck cancer, hepatocellular cancer, hodgkin lymphoma, hypopharyngeal cancer, hypothalamic and visual pathway glioma, intraocular melanoma, kaposi sarcoma, laryngeal cancer, medulloblastoma, medulloepithelioma, melanoma, merkel cell carcinoma, mesothelioma, mouth cancer, multiple endocrine neoplasia syndrome, multiple myeloma, mycosis fungoides, nasal cavity and paranasal sinus cancer, nasopharyngeal cancer, neuroblastoma, non-hodgkin lymphoma, non-small cell lung cancer, oral cancer, oropharyngeal cancer, osteosarcoma, ovarian cancer, ovarian epithelial cancer, ovarian germ cell tumor, ovarian low malignant potential tumor, pancreatic cancer, papillomatosis, paranasal sinus and nasal cavity cancer, parathyroid cancer, penile cancer, pharyngeal cancer, pheochromocytoma, pituitary tumor, pleuropulmonary blastoma, primary central nervous system lymphoma, prostate cancer, rectal cancer, renal cell cancer, retinoblastoma, rhabdomyosarcoma, salivary gland cancer, sézary syndrome, small cell lung cancer, small intestine cancer, soft tissue sarcoma, squamous cell carcinoma, squamous neck cancer, testicular cancer, throat cancer, thymic carcinoma, thymoma, thyroid cancer, urethral cancer, uterine sarcoma, vaginal cancer, vulvar cancer, waldenström macroglobulinemia, and wilms tumor. More preferably, the cancer is a solid cancer, a metastasis, or a relapse thereof. More preferably, said cancer is glioblastoma, pancreatic ductal adenocarcinoma, osteosarcoma, or a brain metastasis of a non-brain primary tumor.

25

The term "cancer antigen" relates to an antigen, preferably a polypeptide, expressed by a cancer cell. Preferably, the cancer antigen is expressed at an at least 5fold, preferably at least 10fold, more preferably at least 25fold, lower rate in non-cancer cells. Preferably, the cancer antigen is not expressed in non-tumor cells of the same tissue in a subject, more preferably is not expressed in non-cancer cells of a subject; thus, the cancer antigen preferably is a cancer specific antigen. More preferably, the cancer antigen is a neoantigen and/or comprises a neoepitope, expressed by cancer cells. Preferably, one or more peptides of the cancer antigen are presented

30

via MHC molecules, more preferably MHC class-I molecules, on the surface of host cells producing said cancer antigen as "cancer epitopes", which preferably are cancer -specific epitopes or, as specified above, cancer neoepitopes. As specified elsewhere herein, the cancer preferably is a solid cancer, i.e. a tumor-forming cancer; thus, the cancer antigen preferably is a tumor antigen, more preferably a tumor-specific antigen, and the cancer epitope preferably is a tumor epitope, more preferably a tumor-specific epitope.

The term "autoimmune T-cell activating antigen" is, in principle, known to the skilled person to relate to any antigen presented by a cell of a subject, the recognition of which causes, aggravates, or contributes to autoimmune disease, preferably T-cell mediated autoimmune disease. T-cell mediated autoimmune diseases are known in the art; preferably, the T-cell mediated autoimmune disease is selected from the list consisting of multiple sclerosis, celiac disease, rheumatoid arthritis, type 1 diabetes mellitus, hypothyroidism, and Addison's disease. As the skilled person understands, identification of autoimmune-reactive T-cells and/or their TCRs as proposed herein preferably is particularly suitable for diagnosing, contributing to diagnosing, and/or predicting T-cell mediated autoimmune disease. The autoimmune-reactive T-cells and/or their TCRs may, however, also be used for generation of regulatory T-cells and, therefore, be used in the treatment of T-cell mediated autoimmune disease. Furthermore, the autoimmune-reactive T-cells and/or their TCRs preferably are used in the identification of new autoimmune T-cell activating antigens.

As used herein, the term "host cell" relates to any cell capable of expressing, and preferably presenting on its surface, a TCR polypeptide as specified herein, preferably encoded by a polynucleotide and/or vector. Preferably, the cell is a bacterial cell, more preferably a cell of a common laboratory bacterial strain known in the art, most preferably an Escherichia strain, in particular an E. coli strain. Also preferably, the host cell is a eukaryotic cell, preferably a yeast cell, e.g. a cell of a strain of baker's yeast, or is an animal cell. More preferably, the host cell is an insect cell or a mammalian cell, in particular a mouse or rat cell. Most preferably, the host cell is a human cell. Preferably, the host cell is a T-cell, more preferably a CD8⁺ T-cell or a CD4⁺ T-cell, more preferably a CD8⁺ T-cell. As the skilled person understands, a CD4 TCR is preferably expressed in a CD8⁺ T-cell, and a CD8 TCR is preferably expressed in a CD4⁺ T-cell.

The terms "identifying a T-cell reactive to cells presenting a T-cell activating antigen" and "identifying a reactive T-cell", as used herein, are used in a broad sense including any and all means and methods of providing information on a reactive T-cell allowing determination of at least the CDR sequences of its TCR. In accordance, the reactive T-cell does not have to, but may, be provided in physical form. Thus, identifying a reactive T-cell may comprise identifying a dataset indicative of a T-cell expressing at least one biomarker as specified elsewhere herein and, optionally, allocating at least the CDR sequences of the TCR of said reactive T-cell. Preferably, said dataset is or was determined by single-cell determination of gene expression, preferably by single-cell RNA sequencing. Thus, step a) of the method of identifying a reactive T-cell may comprise performing single-cell determination of gene expression of T-cells in a sample, wherein expression of at least one of the biomarkers as specified is determined, thereby identifying a reactive T-cell; optionally, at least the CDR sequences of the TCR of said T-cell found to express said at least one biomarker are sequenced. Identifying a reactive T-cell may, however, also comprise physically providing said reactive T-cell. Thus, step a) of the method of identifying a reactive T-cell may comprise determining expression of at least one of the biomarkers as specified on and/or in the T-cell. Thus, expression of surface biomarkers may e.g. be determined by antibody staining, optionally followed by FACS-measurements and/or -sorting. Also preferably, single T-cells are grown clonally and biomarker expression is determined in an aliquot of said clonally grown cells. Other methods of determining biomarker expression in a T-cell, preferably a living T-cells, are known in the art.

Determination of expression of a biomarker may be performed based on the amount of any biomarker gene product deemed appropriate by the skilled person. Thus, determination may comprise determining the amount of RNA, in particular mRNA, and/or polypeptide gene product. Expression may, however, also be determined by measuring expression of a surrogate biomarker, e.g. a reporter gene construct in which the reporter gene is expressed under the control of the promoter of the respective biomarker. Preferably, the determination of expression comprises determining the amount of mRNA and/or polypeptide gene product.

Identifying a reactive T-cell comprises determining expression of at least one biomarker as specified elsewhere herein. Expression of a biomarker may be determined qualitatively, semi-quantitatively, or quantitatively, which terms are in principle known to the skilled person. Qual-

- itative determination may be a binary assessment that the biomarker is expressed or not expressed by a T-cell, e.g. by determining whether the biomarker is expressed above a detection level of an assay. Semiquantitative determination may comprise assorting expression to expression categories, such as low, medium, or high expression. The term quantitative determination
- 5 is understood by the skilled person to include each and every determination providing information on the amount of a biomarker in a cell and all values derived from such an amount by at least one standard mathematical operation, including in particular calculation of a concentration, of a mean, a median, or an average, normalization, and similar calculations.
- 10 Preferably, identifying a reactive T-cell comprises comparing biomarker expression determined in a T-cell to a reference. The term "reference", as used herein, refers to expression of a biomarker in a reference cell, e.g. an amount of biomarker in a reference cell. Preferably, a reference is a threshold value (e.g., an amount or ratio of amounts) for a gene product. The reference may, however, also be a value derived from an amount by any mathematical deemed appropriate
- 15 by the skilled person, in particular normalization. In accordance with the aforementioned method, a reference is, preferably, a reference obtained from a sample of T-cells known to be reactive T-cells. In such a case, a value for the biomarker gene product found in a sample being essentially identical to said reference is indicative for a reactive T-cell. Also preferably, the reference is from a sample of T-cells known not be reactive. In such a case, a value for the
- 20 biomarker gene product found in the T-cell to be increased with respect to the reference is indicative for the T-cell being reactive. The same applies mutatis mutandis for a calculated reference, most preferably the average or median, for the relative or absolute value of the biomarker gene product(s) of a population of non-stimulated T-cells. As the skilled person understands, only a small percentage of T-cells of any given natural population of T-cells will be
- 25 reactive at a time. In accordance, the above description for a population of T-cells known not to be activated may be applied mutatis mutandis to a natural population of T-cells of which the activation status is unknown; thus the reference may be a natural sample of T-cells of which reactivity status is unknown. In such a case, a value for the biomarker gene product found in the T-cell to be increased with respect to the reference is indicative for the T-cell being reactive.
- 30 How to calculate a suitable reference value, preferably, the average or median, is well known in the art. The population of non-stimulated T-cells referred to before shall comprise a plurality of T-cells, preferably at least 10, more preferably at least 100, still more preferably at least 1,000, most preferably at least 10,000, non-stimulated T-cells. The value for a biomarker gene

product of T-cell of interest and the reference values are essentially identical if the corresponding values are essentially identical. Essentially identical means that the difference between two values is, preferably, not significant and shall be characterized in that the values are within at least the interval between 1st and 99th percentile, 5th and 95th percentile, 10th and 90th percentile, 20th and 80th percentile, 30th and 70th percentile, 40th and 60th percentile of the reference value, preferably, the 50th, 60th, 70th, 80th, 90th or 95th percentile of the reference value. Statistical tests for determining whether two amounts are essentially identical are well known in the art. An observed difference for two values, on the other hand, shall preferably be statistically significant. A difference in the relative or absolute value is, preferably, significant outside of the interval between 45th and 55th percentile, 40th and 60th percentile, 30th and 70th percentile, 20th and 80th percentile, 10th and 90th percentile, 5th and 95th percentile, 1st and 99th percentile of the reference value. Preferably, the reference(s) are stored in a suitable data storage medium such as a database and are, thus, also available for future assessments.

Identifying a reactive T-cell comprises determining expression of at least one of CCL4, CCL4L2, CCL3, CCL3L1, and CXCL13, preferably of at least one of CCL4, CCL4L2, CCL3, and CCL3L1. Thus, the method of identifying a reactive T-cell preferably comprises determining expression of at least one biomarker selected from the list consisting of CCL4, CCL4L2, CCL3, CCL3L1, and CXCL13. Thus, the method of identifying a reactive T-cell preferably comprises determining expression of at least one biomarker selected from Table 1 herein below. The aforesaid biomarkers are biomarkers of the "core signature", i.e. each biomarker alone or any combination thereof is indicative of a reactive T-cell. The aforesaid biomarkers are, in principle, known to the skilled person and their amino acid sequences and sequences of encoding polynucleotides are available from public databases. "CCL4" is also known as "Chemokine (C-C motif) ligand 4" and the amino acid sequence of human CCL4 is available e.g. from Genbank Acc No. NP_996890.1. "CCL4L2" is also known as "C-C motif chemokine 4-like 2" and the amino acid sequence of human CCL4L2 is available e.g. from Genbank Acc No. NP_001278397.1. "CCL3" is also known as "Chemokine (C-C motif) ligand 3" and may also be referred to as macrophage inflammatory protein 1-alpha (MIP-1-alpha); the amino acid sequence of human CCL3 is available e.g. from Genbank Acc No. NP_002974.1. "CCL3L1" is also known as Chemokine (C-C motif) ligand 3-like 1", and the amino acid sequence of human CCL3L1 is available e.g. from Genbank Acc No. NP_066286.1. "CXCL13" is also known under the designations "B lymphocyte chemoattractant" and "B cell-attracting chemokine 1", and

the amino acid sequence of human CXCL13 is available e.g. from Genbank Acc No. NP_006410.1. Preferably, expression of at least one of CCL4, CCL4L2, CCL3, CCL3L1, and CXCL13 is indicative of a reactive T-cell. More preferably, expression of at least two, more preferably at least three, most preferably all four of the aforesaid biomarkers is indicative of a reactive T-cell.

Thus, in a preferred embodiment, identifying a reactive T-cell comprises determining expression of a biomarker combination comprising, preferably consisting of, CCL3L1; CCL3L1 + CCL4; CCL3L1 + CCL4L2; CCL3L1 + CCL3; CCL3L1 + CXCL13; CCL3L1 + CCL4 + CCL4L2; CCL3L1 + CCL4 + CCL3; CCL3L1 + CCL4 + CXCL13; CCL3L1 + CCL4L2 + CCL3; CCL3L1 + CCL4L2 + CXCL13; CCL3L1 + CCL3 + CXCL13; CCL3L1 + CCL4 + CCL4L2 + CCL3; CCL3L1 + CCL4 + CCL4L2 + CXCL13; CCL3L1 + CCL4 + CCL3 + CXCL13; CCL3L1 + CCL4L2 + CCL3 + CXCL13; CCL3L1 + CCL4 + CCL4L2 + CCL3 + CXCL13; CCL4; CCL4 + CCL4L2; CCL4 + CCL3; CCL4 + CXCL13; CCL4 + CCL4L2 + CCL3; CCL4 + CCL4L2 + CXCL13; CCL4 + CCL3 + CXCL13; CCL4 + CCL4L2 + CCL3 + CXCL13; CCL4L2; CCL4L2 + CCL3; CCL4L2 + CXCL13; CCL4L2 + CCL3 + CXCL13; CCL3; CCL3 + CXCL13; or CXCL13.

Preferably, the method of identifying a reactive T-cell referred to herein comprises further determining expression of at least one biomarker selected from the list consisting of IFNG, HAVCR2, FNBP1, CSRNP1, SPRY1, RHOH, FOXN2, HIF1A, TOB1, RILPL2, CD8B, GABARAPL1, TNFSF14, EGR1, EGR2, TAGAP, TNFSF9, ANXA1, MAP3K8, PIK3R1, DUSP2, DUSP4, DUSP6, CLIC3, RASGEF1B, LAG3, XCL2, NR4A2, DNAJB6, NFKBID, MCL1, EVI2A, SLC7A5, H3F3B, NR4A3, REL, IRF4, CST7, ATF3, TNF, GPR171, BCL2A1, ITGA1, TNFAIP3, NR4A1, RUNX3, HERPUD2, FASLG, CBLB, PTGER4, SLA, XCL1, BHLHE40, LYST, KLRD1, ZNF682, CTSW, SLC2A3, NLRP3, SCML4, VSIR, LINC01871, and ZFP36L1. Thus, the method of identifying a reactive T-cell preferably comprises determining expression of at least one biomarker selected from Table 2 herein below. The biomarkers are biomarkers of the "accessory 1 signature", i.e. each biomarker of Table 2, alone or in combination with at least one further biomarker of Table 2, is indicative of a reactive T-cell if determined in combination with at least one biomarker of Table 1. Thus, preferably, expression of at least one biomarker of Table 2 in addition to at least one of CCL4, CCL4L2, CCL3, CCL3L1, and CXCL13 is indicative of a reactive T-cell.

Preferably, the method of identifying a reactive T-cell referred to herein comprises further determining expression of at least one biomarker selected from the list consisting of CCL5, GZMH, CLEC2B, GZMA, CD69, GZMK, and CRTAM. Thus, the method of identifying a reactive T-cell preferably comprises determining expression of at least one biomarker selected from Table 3 herein below. The biomarkers are biomarkers of the "accessory 2 signature", i.e. each biomarker of Table 3, alone or in combination with at least one further biomarker of Table 2 or Table 3, is indicative of a reactive T-cell if determined in combination with at least one biomarker of Table 1. Thus, preferably, expression of at least one biomarker of Table 3 in addition to at least one of CCL4, CCL4L2, CCL3, CCL3L1, and CXCL13 is indicative of a reactive T-cell.

In view of the above, all biomarkers of Tables 1 to 3, when expressed in a T-cell, are indicative of a reactive and/or can contribute to identification of a reactive T-cell. Preferably, the method further comprises determination of at least one exclusion biomarker, i.e. a biomarker which, when expressed, is indicative that the T-cell is non-reactive: Preferably, the method of identifying a reactive T-cell referred to herein comprises further determining expression of at least one biomarker selected from the list consisting of GNLY and FGFBP2 (Table 4), wherein expression of at least one of said biomarkers is indicative of a non-reactive T-cell. Thus, the biomarker(s) GNLY and/or FGFBP2 may be used as an exclusion biomarker. In a preferred embodiment, FOXP3 may be used as a (further) exclusion biomarker.

In a preferred embodiment, identifying a reactive T-cell comprises determining expression of at least one of CCL3L1, LAG3, GABARAPL1, CBLB, SLA, KLRD1, and CLEC2B. Thus, the method of identifying a reactive T-cell preferably comprises determining expression of at least one biomarker selected from the list consisting of CCL3L1, LAG3, GABARAPL1, CBLB, SLA, KLRD1, and CLEC2B. Thus, the method of identifying a reactive T-cell preferably comprises determining expression of at least one biomarker selected from Table 5 herein below. The aforesaid biomarkers are biomarkers of the "core-2 signature", i.e. each biomarker alone or any combination thereof is indicative of a reactive T-cell, and may collectively also be referred to herein as "7 alternative core genes". The aforesaid biomarkers are, in principle, known to the skilled person and their amino acid sequences and sequences of encoding polynucleotides are available from public databases. CCL3L1 was described herein above. "LAG3" is also known as lymphocyte activating 3 and the amino acid sequence of human LAG3 is available

e.g. from Genbank Acc No. NP_002277.4. "GABARAPL1" is also known as GABA type A receptor associated protein like 1 and the amino acid sequence of human GABARAPL1 is available e.g. from Genbank Acc No. NP_001350527.1. "CBLB" is also known as Cbl proto-oncogene B and the amino acid sequence of human CBLB is available e.g. from Genbank Acc No. NP_001308715.1. "SLA" is also known as, Src like adaptor and the amino acid sequence of human SLA is available e.g. from Genbank Acc No. NP_001039021.1. "KLRD1" is also known as killer cell lectin like receptor D1 and the amino acid sequence of human KLRD1 is available e.g. from Genbank Acc No. NP_001107868.2. "CLEC2B" is also known as C-type lectin domain family 2 member B and the amino acid sequence of human CLEC2B is available e.g. from Genbank Acc No. NP_005118.2. Preferably, expression of at least one of CCL3L1, LAG3, GABARAPL1, CBLB, SLA, KLRD1, and CLEC2B is indicative of a reactive T-cell. More preferably, expression of at least two, more preferably at least three, even more preferably at least four, even more preferably at least five, still more preferably at least six, most preferably all seven of the aforesaid biomarkers is indicative of a reactive T-cell.

15

Thus, in a preferred embodiment, identifying a reactive T-cell comprises determining expression of a biomarker combination comprising, preferably consisting of, CCL3L1; CCL3L1 + LAG3; CCL3L1 + GABARAPL1; CCL3L1 + CBLB; CCL3L1 + SLA; CCL3L1 + KLRD1; CCL3L1 + CLEC2B; CCL3L1 + LAG3 + GABARAPL1; CCL3L1 + LAG3 + CBLB; CCL3L1 + LAG3 + SLA; CCL3L1 + LAG3 + KLRD1; CCL3L1 + LAG3 + CLEC2B; CCL3L1 + GABARAPL1 + CBLB; CCL3L1 + GABARAPL1 + SLA; CCL3L1 + GABARAPL1 + KLRD1; CCL3L1 + GABARAPL1 + CLEC2B; CCL3L1 + CBLB + SLA; CCL3L1 + CBLB + KLRD1; CCL3L1 + CBLB + CLEC2B; CCL3L1 + SLA + KLRD1; CCL3L1 + SLA + CLEC2B; CCL3L1 + KLRD1 + CLEC2B; CCL3L1 + LAG3 + GABARAPL1 + CBLB; CCL3L1 + LAG3 + GABARAPL1 + SLA; CCL3L1 + LAG3 + GABARAPL1 + KLRD1; CCL3L1 + LAG3 + GABARAPL1 + CLEC2B; CCL3L1 + LAG3 + CBLB + SLA; CCL3L1 + LAG3 + CBLB + KLRD1; CCL3L1 + LAG3 + CBLB + CLEC2B; CCL3L1 + LAG3 + SLA + KLRD1; CCL3L1 + LAG3 + SLA + CLEC2B; CCL3L1 + LAG3 + KLRD1 + CLEC2B; CCL3L1 + GABARAPL1 + CBLB + SLA; CCL3L1 + GABARAPL1 + CBLB + KLRD1; CCL3L1 + GABARAPL1 + CBLB + CLEC2B; CCL3L1 + GABARAPL1 + SLA + KLRD1; CCL3L1 + GABARAPL1 + SLA + CLEC2B; CCL3L1 + GABARAPL1 + KLRD1 + CLEC2B; CCL3L1 + CBLB + SLA + KLRD1; CCL3L1 + CBLB + SLA + CLEC2B; CCL3L1 + CBLB + KLRD1 + CLEC2B; CCL3L1 + SLA + KLRD1 + CLEC2B; CCL3L1 + LAG3 + GABARAPL1 +

30

CBLB + SLA; CCL3L1 + LAG3 + GABARAPL1 + CBLB + KLRD1; CCL3L1 + LAG3 + GABARAPL1 + CBLB + CLEC2B; CCL3L1 + LAG3 + GABARAPL1 + SLA + KLRD1; CCL3L1 + LAG3 + GABARAPL1 + SLA + CLEC2B; CCL3L1 + LAG3 + GABARAPL1 + KLRD1 + CLEC2B; CCL3L1 + LAG3 + CBLB + SLA + KLRD1; CCL3L1 + LAG3 + CBLB + SLA + CLEC2B; CCL3L1 + LAG3 + CBLB + KLRD1 + CLEC2B; CCL3L1 + LAG3 + SLA + KLRD1 + CLEC2B; CCL3L1 + GABARAPL1 + CBLB + SLA + KLRD1; CCL3L1 + GABARAPL1 + CBLB + SLA + CLEC2B; CCL3L1 + GABARAPL1 + CBLB + KLRD1 + CLEC2B; CCL3L1 + GABARAPL1 + SLA + KLRD1 + CLEC2B; CCL3L1 + CBLB + SLA + KLRD1 + CLEC2B; CCL3L1 + LAG3 + GABARAPL1 + CBLB + SLA + KLRD1; CCL3L1 + LAG3 + GABARAPL1 + CBLB + SLA + CLEC2B; CCL3L1 + LAG3 + GABARAPL1 + CBLB + KLRD1 + CLEC2B; CCL3L1 + LAG3 + CBLB + SLA + KLRD1 + CLEC2B; CCL3L1 + GABARAPL1 + CBLB + SLA + KLRD1 + CLEC2B; CCL3L1 + LAG3 + GABARAPL1 + CBLB + SLA + KLRD1 + CLEC2B; LAG3; LAG3 + GABARAPL1; LAG3 + CBLB; LAG3 + SLA; LAG3 + KLRD1; LAG3 + CLEC2B; LAG3 + GABARAPL1 + CBLB; LAG3 + GABARAPL1 + SLA; LAG3 + GABARAPL1 + KLRD1; LAG3 + GABARAPL1 + CLEC2B; LAG3 + CBLB + SLA; LAG3 + CBLB + KLRD1; LAG3 + CBLB + CLEC2B; LAG3 + SLA + KLRD1; LAG3 + SLA + CLEC2B; LAG3 + KLRD1 + CLEC2B; LAG3 + GABARAPL1 + CBLB + SLA; LAG3 + GABARAPL1 + CBLB + KLRD1; LAG3 + GABARAPL1 + CBLB + CLEC2B; LAG3 + GABARAPL1 + SLA + KLRD1; LAG3 + GABARAPL1 + SLA + CLEC2B; LAG3 + GABARAPL1 + KLRD1 + CLEC2B; LAG3 + CBLB + SLA + KLRD1; LAG3 + CBLB + SLA + CLEC2B; LAG3 + CBLB + KLRD1 + CLEC2B; LAG3 + SLA + KLRD1 + CLEC2B; LAG3 + GABARAPL1 + CBLB + SLA + KLRD1; LAG3 + GABARAPL1 + CBLB + SLA + CLEC2B; LAG3 + GABARAPL1 + CBLB + KLRD1 + CLEC2B; LAG3 + GABARAPL1 + SLA + KLRD1 + CLEC2B; LAG3 + CBLB + SLA + KLRD1 + CLEC2B; LAG3 + GABARAPL1 + CBLB + SLA + KLRD1 + CLEC2B; GABARAPL1; GABARAPL1 + CBLB; GABARAPL1 + SLA; GABARAPL1 + KLRD1; GABARAPL1 + CLEC2B; GABARAPL1 + CBLB + SLA; GABARAPL1 + CBLB + KLRD1; GABARAPL1 + CBLB + CLEC2B; GABARAPL1 + SLA + KLRD1; GABARAPL1 + SLA + CLEC2B; GABARAPL1 + KLRD1 + CLEC2B; GABARAPL1 + CBLB + SLA + KLRD1; GABARAPL1 + CBLB + SLA + CLEC2B; GABARAPL1 + CBLB + KLRD1 + CLEC2B; GABARAPL1 + SLA + KLRD1 + CLEC2B; GABARAPL1 + CBLB + SLA + KLRD1 + CLEC2B; CBLB; CBLB + SLA; CBLB + KLRD1; CBLB + CLEC2B; CBLB + SLA + KLRD1; CBLB + SLA + CLEC2B; CBLB +

KLRD1 + CLEC2B; CBLB + SLA + KLRD1 + CLEC2B; SLA; SLA + KLRD1; SLA + CLEC2B; SLA + KLRD1 + CLEC2B; KLRD1; KLRD1 + CLEC2B; or CLEC2B.

In a preferred embodiment, identifying a reactive T-cell comprises determining expression of a biomarker combination comprising, preferably consisting of, at least one biomarker combination ("signature") disclosed in any one of Tables 7 to 10 or any combination thereof, preferably at least one biomarker combination disclosed in Table 7. In a further preferred embodiment, identifying a reactive T-cell comprises determining expression of a biomarker combination comprising, preferably consisting of, at least one biomarker combination disclosed in Table 7 or 8 and the cancer is a non-primary brain metastasis. In a further preferred embodiment, identifying a reactive T-cell comprises determining expression of a biomarker combination comprising, preferably consisting of, at least one biomarker combination disclosed in Table 7 or 9 and the cancer is a lung cancer. In a further preferred embodiment, identifying a reactive T-cell comprises determining expression of a biomarker combination comprising, preferably consisting of, all biomarkers of Table 5 + CD8B; all biomarkers of Tables 1 and 2 + CD8B; all biomarkers of Table 1 + CD8B; all biomarkers of Tables 5 and 6 + CD8B, and the cancer is a glioma.

The skilled person is aware that the biomarkers referred to herein may be expressed in a plurality of isoforms, from different alleles, and/or may be expressed as precursor forms which may be further processed in the cell, e.g. during intracellular trafficking and/or secretion. Also, the skilled person is aware that subjects from non-human species will preferably express homologues of the specific sequences indicated herein above, which may preferably be identified by sequence alignment and/or search algorithms based thereon, such as the BLAST algorithm, and appropriate databases, preferably publicly available databases. Preferably, the amino acid sequence of a biomarker as specified is at least 50%, more preferably 75%, still more preferably 85%, even more preferably at least 95%, even more preferably at least 98%, most preferably at least 99%, identical to a specific biomarker sequence as referred to herein.

The term "subject", as used herein, relates to an animal, preferably a vertebrate, more preferably a mammal, preferably to a livestock, like a cattle, a horse, a pig, a sheep, or a goat, to a companion animal, such as a cat or a dog, or to a laboratory animal, like a rat, mouse, or guinea pig. Preferably, the mammal is a primate, more preferably a monkey, most preferably a human.

Preferably, the subject is suffering from cancer, in particular in case of the method of identifying a T-cell reactive to cancer cells of a subject. It is, however, also envisaged that the subject is an apparently healthy subject, preferably at least 50 years, more preferably at least 60 years, more preferably at least 70 years, even more preferably at least 80 years of age.

5

The term "sample" refers to a sample of separated cells or to a sample from a tissue or an organ, preferably from a tumor. Thus, the sample preferably comprises or is assumed to comprise cancer recognizing lymphocytes, preferably T-cells. More preferably, the sample comprises or is assumed to comprise tumor-infiltrating lymphocytes (TILs). Also preferably, the sample
10 comprises cancer cells, more preferably tumor cells. Thus, the sample preferably comprises TILs and cancer cells, preferably is a tumor sample. The sample may, however also be a sample of non-cancer tissue, preferably of cancer-adjacent tissue, or a sample of peripheral blood monocytes (PBMCs). As is known to the skilled person, tissue or organ samples may be obtained from any tissue or organ by, e.g., biopsy, surgery, or any other method deemed appropriate by
15 the skilled person. Separated cells may be obtained from the body fluids, such as lymph, blood, plasma, serum, liquor and other, or from the tissues or organs by separating techniques such as centrifugation or cell sorting. Preferably, the sample is a tissue or body fluid sample which comprises cells. Preferably the sample is a sample of a body fluid, preferably a blood sample. The body fluid sample can be obtained from the subject by routine techniques which are well
20 known to the person skilled in the art, e.g., venous or arterial puncture, lavage, or any other method deemed appropriate by the skilled person.

Advantageously, it was found in the work underlying the present invention that using the biomarkers CCL4, CCL4L2, CCL3, CCL3L1, and/or CXCL13, optionally including further bi-
25 omarkers, allows identification of T-cells which comprise TCRs which are reactive to antigens presented by cells and, therefore, particularly suitable for providing, either by cultivation or by expressing the respective TCR in a T-cell, T-cells recognizing e.g. cancer cells, e.g. for cellular therapy of cancer.

30 The definitions made above apply mutatis mutandis to the following. Additional definitions and explanations made further below also apply for all embodiments described in this specification mutatis mutandis.

The present invention also relates to a method of identifying a T-cell reactive to cells of a subject presenting a T-cell activating antigen (reactive T-cell), comprising

(a) determining expression of at least one of KLRD1 and LAG3 in T-cells from a sample of said subject; and

- 5 (b) identifying a reactive T-cell based on the determination of step (a), preferably wherein said T-cell activating antigen is a cancer antigen or an autoimmune T-cell antigen, more preferably is a cancer antigen.

10 The present invention further relates to a method of identifying a TCR binding to an activating antigen presented on a cell, preferably a cancer cell, of a subject, said method comprising

(A) identifying a reactive T-cell according to the method of identifying reactive T-cells,
(B) providing the amino acid sequences of at least the complementarity determining regions (CDRs) of the TCR of the reactive T-cell identified in step (A); and, hereby,
(C) identifying a TCR binding to an activating antigen presented on a cell.

15

The method of identifying a TCR, preferably, is an in vitro method. The method may comprise further steps in addition to those related to herein above. For example, further steps may relate, e.g., to determining further nucleic acid or amino acid sequences, or determining CD8 and/or CD4 expression by said reactive T-cell. Moreover, one or more of said steps may be performed
20 or assisted by automated equipment.

20

The term "providing a sequence", such as an amino acid sequence and/or a nucleic acid sequence, is used herein in a broad sense including any and all means and methods of providing information on said sequence or making said sequence information accessible. Thus, the sequence may be provided as a sequence information, preferably tangibly embedded on a data
25 carrier. The sequence may, however, also be provided in the form of a molecule comprising said sequence, preferably as a TCR comprising TCR alpha and beta chains comprising said sequences, more preferably as a host cell comprising the same. As the skilled person understands, if the aforesaid host cell is provided, the sequence information can be provided by standard methods known to the skilled person, e.g. nucleic acid sequencing of the TCR expressed by
30 said host cell or of parts thereof.

30

The term "identifying a TCR" is used herein in a broad sense including any and all means and methods of providing information on a TCR allowing determination of at least its CDR sequences. In accordance, the TCR does not have to, but may, be provided in physical form. Thus, identifying a TCR may comprise providing at least the CDR sequences of the TCR or of a polynucleotide encoding at least said CDRs. Preferably, said sequences are or were determined by single-cell determination of gene expression, preferably by single-cell RNA sequencing, preferably as specified herein above. Identifying a TCR may, however, also comprise physically providing said TCR, e.g. by providing a host cell, preferably a T-cell, expressing said TCR, or by providing at least one polynucleotide encoding at least the CDRs of the TCR polypeptides identified. As will be understood, in case the TCR is provided in the context of a self-replicating entity such as a host cell, it may not be necessary to provide the amino acid of at least the CDRs of the TCR and/or the nucleic acid complex of a polynucleotide encoding the same.

Preferably, the method of identifying a TCR binding to an activating antigen further comprises step B1) expressing a TCR comprising at least the CDRs determined in step B) in a host cell, preferably a T-cell. More preferably, said method comprises further step B1) expressing a TCR comprising at least the CDRs determined in step B) in a host cell, preferably a T-cell, i.e. preferably comprises expressing a TCR comprising at least the CDRs determined in step B) and at least one accessory TCR polypeptide in a host cell.

The term "TCR comprising at least the CDRs" as specified, as used herein, relates to a TCR in which at least the CDRs are those as determined in step B), while the residual sequences of the TCR polypeptides may be sequences of one or more different alpha and beta or gamma and delta chains, e.g. heterologous sequences. More preferably, the variable regions of the TCR molecules are provided in step B) and are expressed as parts of the TCR polypeptides in step B1). It is, however, also envisaged that sequences of further fragments of the TCR polypeptides or the complete TCR polypeptides are provided in step B), and are optionally expressed in step B1). As the skilled person understands, it is also possible to provide longer sequences in step B) than are expressed in step B1); e.g., preferably, the amino acid sequence of the variable regions of the TCR polypeptides may be provided in step B), while only the CDRs thereof are expressed, in the context e.g. of heterologous TCR polypeptides, in step B1); or, the amino acid sequences of the variable regions of the TCR polypeptides may be provided in step B), and the

amino acid sequence of the antigen binding region, including the CDRs, may be expressed, in the context e.g. of heterologous TCR polypeptides, in step B1). If not otherwise indicated, the TCR polypeptides preferably are expressed as complete molecules, i.e. each comprising a trans-membrane region, a constant region, a joining region, and a variable region.

5

Preferably, the method of identifying a TCR binding to a cell presenting a T-cell activating antigen comprises further step B2) determining binding of the TCR expressed in step B1) to a cell presenting a T-cell activating antigen, preferably a cancer antigen, complexed in a major histocompatibility complex (MHC), preferably MHC class I, molecule. Methods of determining binding of a TCR, preferably comprised in a TCR, to a T-cell activating antigen complexed in a major histocompatibility complex (MHC) molecule are known in the art and include, preferably, determining binding of a T-cell activating antigen complexed an MHC molecule carrying a detectable label to the TCR which may e.g. be expressed on the surface of a host cell. A well-known example of such a method is a tetramer assay, preferably using a soluble tetrameric MHC molecule complexed with a T-cell activating antigen.

Preferably, the method identifying a TCR binding to a T-cell activating antigen comprises further step B3) determining recognition of cells presenting said T-cell activating antigen by the TCR expressed in step B1). Assays for determining such recognition are known in the art and include in particular binding assays, activation assays, and lysis assays. In all of these assays, preferably cells presenting T-cell activating antigen are co-incubated with host cells such as T-cells expressing a TCR comprising at least the CDRs as specified. In a binding assay, it is determined whether the cancer cells and the aforesaid host cells bind to each other, preferably to form an immunological synapse including at least an MHC molecule of the cell presenting a T-cell activating antigen and the TCR. In an activation assay, the host cell, preferably the T-cell, expressing a TCR comprising at least the CDRs as specified, is tested after said co-incubation for biomarkers of immunological activation, e.g. interferon-gamma production. In a lysis assay, it is determined whether the host cells, preferably the T-cells, expressing a TCR comprising at least the CDRs as specified, lysed at least a fraction of the cells presenting the T-cell activating antigen during said co-incubation.

Preferably, the method of identifying a TCR binding to an activating antigen comprises further step B4) producing a soluble TCR comprising at least the CDRs determined in step B) and

determining binding of said soluble TCR to a cancer cell and/or to a cancer antigen complexed in a major histocompatibility complex (MHC), preferably MHC class I, molecule. Soluble TCRs have been described herein above. Preferably, soluble TCRs carrying a detectable label are used in step B4); thus, binding of a such labeled soluble TCR may e.g. be detected by fluorescence-activated cell sorting equipment.

Preferably, in the method of identifying a TCR, expression of said at least one biomarker is determined by single-cell determination of gene expression, preferably of at least 100 T-cells, more preferably at least 1000 T-cells. In such case, the amino acid sequences of at least the complementarity determining regions (CDRs) of the TCR of the reactive T-cell of step (B) may be provided as part of the single-cell determination of gene expression, i.e. the mRNAs encoding said CDRs may be sequenced as part of said single-cell determination of gene expression. Preferably, corresponding sequences are pre-amplified before single-cell determination of gene expression. The mRNAs encoding said CDRs may, however, also be determined in separate sequencing steps, preferably by using appropriate barcoding methods. Also in the aforesaid case, the method may comprise further step (B*1) clustering the T-cells based on said gene expression data including the amino acid sequences of said at least CDR sequences and further step (B*2) selecting the TCR or TCRs being clustered at increased relative frequency in clusters expressing said at least one biomarker compared to clusters not expressing said at least one biomarker.

The present invention also relates to a method of providing a T-cell recognizing a cell presenting a T-cell activating antigen, preferably a cancer cell, said method comprising

- (i) identifying a TCR binding to a cell presenting a T-cell activating antigen according to the method according to the present invention,
- (ii) expressing a TCR comprising at least the complementarity determining regions (CDRs) of the TCR of step (I) in a T-cell, and, thereby,
- (iii) providing a T-cell recognizing a cell presenting a T-cell activating antigen, preferably a cancer cell.

The method of providing a T-cell, preferably, is an in vitro method, of which one or more steps may be performed or assisted by automated equipment.

The method may comprise further steps in addition to those related to herein above. For example, further steps may relate, e.g., to cloning at least polynucleotides encoding the CDRs of the TCR of step (i) into a TCR alpha, beta, gamma, or delta chain backbone, or cloning polynucleotides encoding the variable regions of the TCR polypeptides of step (i) into TCR alpha and beta or a TCR gamma and delta, chain backbones, preferably on at least one expression vector; or cloning polynucleotides encoding TCR polynucleotides into one or more expression vectors. As will be understood by the skilled person, CDRs and/or a variable region of a TCR alpha chain will preferably be cloned into a TCR alpha chain backbone; and CDRs and/or a variable region of a TCR beta chain will preferably be cloned into a TCR beta chain backbone. The aforesaid applies mutatis mutandis to gamma and delta chains. The method may also comprise the further step of expanding, preferably clonally expanding, the T-cell recognizing a cancer cell to provide a preparation of cells T-cell recognizing cancer cells. It will thus be appreciated that the T-cell recognizing cancer cells may be the T-cell identified in step (i) or a clonal derivative (i.e. a daughter cell) thereof.

15

Preferably, the method of providing a T-cell further comprises a step of testing reactivity of the T-cell of step (ii) to cells presenting an activating agent, e.g. a cancer cells. The term "testing reactivity of a T-cell", as used herein, includes each and every method deemed suitable by the skilled person to determine whether the T-cell as specified is reactive. Preferred methods for testing reactivity have been described herein above, e.g. determining binding, activation of T cells, and/or lysis of cells presenting an activating antigen, e.g. cancer cells.

20

The present invention further relates to a reactive T-cell identified by the method of identifying a T-cell reactive to cells presenting a T-cell activating antigen, preferably cancer cells, as specified herein above and/or obtained or obtainable by the method of providing a T-cell recognizing a cells presenting a T-cell activating antigen as specified herein above, preferably comprising a T-cell receptor comprising an amino acid sequence of SEQ ID NO:1 and/or SEQ ID NO:2, preferably encoded by a polynucleotide comprising SEQ ID NO:3 and/or 4, respectively, for use in medicine, in particular for use in treating and/or preventing cancer or autoimmune disease in a subject.

30

The terms "treating" and "treatment" refer to an amelioration of the diseases or disorders referred to herein or the symptoms accompanied therewith to a significant extent. Said treating

as used herein also includes an entire restoration of health with respect to the diseases or disorders referred to herein. It is to be understood that treating, as the term is used herein, may not be effective in all subjects to be treated. However, the term shall require that, preferably, a statistically significant portion of subjects suffering from a disease or disorder referred to herein
5 can be successfully treated. Whether a portion is statistically significant can be determined without further ado by the person skilled in the art using various well known statistic evaluation tools, e.g., determination of confidence intervals, p-value determination, Student's t-test, Mann-Whitney test etc. Preferred confidence intervals are at least 90%, at least 95%, at least 97%, at least 98% or at least 99 %. The p-values are, preferably, 0.1, 0.05, 0.01, 0.005, or 0.0001. Preferably, the treatment shall be effective for at least 10%, at least 20% at least 50% at least 60%,
10 at least 70%, at least 80%, or at least 90% of the subjects of a given cohort or population. Preferably, treating cancer is reducing tumor burden in a subject. As will be understood by the skilled person, effectiveness of treatment of e.g. cancer is dependent on a variety of factors including, e.g. cancer stage and cancer type.

15

The term "preventing" refers to retaining health with respect to the diseases or disorders referred to herein for a certain period of time in a subject. It will be understood that the said period of time may be dependent on the amount of the drug compound which has been administered and individual factors of the subject discussed elsewhere in this specification. It is to be understood
20 that prevention may not be effective in all subjects treated with the compound according to the present invention. However, the term requires that, preferably, a statistically significant portion of subjects of a cohort or population are effectively prevented from suffering from a disease or disorder referred to herein or its accompanying symptoms. Preferably, a cohort or population of subjects is envisaged in this context which normally, i.e. without preventive measures according to the present invention, would develop a disease or disorder as referred to herein.
25 Whether a portion is statistically significant can be determined without further ado by the person skilled in the art using various well known statistic evaluation tools discussed elsewhere in this specification.

30 The present invention also relates to a pharmaceutical composition comprising a reactive T-cell identified by the method as specified herein above and/or obtained or obtainable by the method of providing a T-cell recognizing a cell presenting a activating antigen as specified herein

above, preferably comprising a T-cell receptor comprising an amino acid sequence of SEQ ID NO:1 and/or SEQ ID NO:2.

5 The term "pharmaceutical composition", as used herein, relates to a composition comprising the compound or compounds, including host cells, in particular T-cells, as specified herein in a pharmaceutically acceptable form and a pharmaceutically acceptable carrier. The compounds and/or excipients can be formulated as pharmaceutically acceptable salts. Acceptable salts comprise acetate, methylester, HCl, sulfate, chloride and the like. The pharmaceutical compositions are, preferably, administered topically or systemically, preferably intravenously or intratumorally.

10 ally. The compounds can be administered in combination with other drugs either in a common pharmaceutical composition or as separated pharmaceutical compositions wherein said separated pharmaceutical compositions may be provided in form of a kit of parts. In particular, co-administration of adjuvants may be envisaged.

15 The compounds are, preferably, administered in conventional dosage forms prepared by combining the host cells or drugs with standard pharmaceutical carriers according to conventional procedures. These procedures may involve mixing, dispersing, or dissolving the ingredients as appropriate to the desired preparation. It will be appreciated that the form and character of the pharmaceutically acceptable carrier or diluent is dictated by the amount of active ingredient

20 with which it is to be combined, the route of administration and other well-known variables.

The carrier(s) must be acceptable in the sense of being compatible with the other ingredients of the formulation and being not deleterious to the recipient thereof. The pharmaceutical carrier employed may be, for example, either a solid, a gel or, preferably a liquid. Exemplary of liquid

25 carriers are phosphate buffered saline solution, water, emulsions, various types of wetting agents, sterile solutions and the like. Suitable carriers comprise those mentioned above and others well known in the art, see, e.g., Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pennsylvania. The diluent(s) is/are preferably selected so as not to affect the biological activity of the T-cells and potential further pharmaceutically active ingredients. Examples of such diluents are distilled water, physiological saline, Ringer's solutions, dextrose

30 solution, and Hank's solution. In addition, the pharmaceutical composition or formulation may also include other carriers, adjuvants, or nontoxic, nontherapeutic, nonimmunogenic stabilizers and the like.

A therapeutically effective dose refers to an amount of the compounds to be used in a pharmaceutical composition of the present invention which prevents, ameliorates or treats a condition referred to herein. Therapeutic efficacy and toxicity of compounds can be determined by standard pharmaceutical procedures in cell culture or in experimental animals, e.g., by determining
5 the ED₅₀ (the dose therapeutically effective in 50% of the population) and/or the LD₅₀ (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD₅₀/ED₅₀.

The dosage regimen will be determined by the attending physician, preferably taking into account relevant clinical factors and, preferably, in accordance with any one of the methods described elsewhere herein. As is well known in the medical arts, a dosage for any one patient may depend upon many factors, including the patient's size, body surface area, age, the particular compound to be administered, sex, time and route of administration, general health, and other drugs being administered concurrently. Progress can be monitored by periodic assessment. A typical dose can be, for example, in the range of 10⁴ to 10⁹ host cells; however, doses below or above this exemplary range are envisioned, especially considering the aforementioned factors. The pharmaceutical compositions and formulations referred to herein are administered at least once in order to treat or prevent a disease or condition recited in this specification. However, the said pharmaceutical compositions may be administered more than one time, for
10
15
20 example, preferably from one to four times, more preferably two or three times.

The present invention also relates to a polynucleotide encoding at least one TCR binding to an activating antigen provided or identifiable according to the method of identifying a TCR binding to an activating antigen as specified herein.

25

The term "polynucleotide" is known to the skilled person. As used herein, the term includes nucleic acid molecules comprising or consisting of a nucleic acid sequence or nucleic acid sequences as specified herein. The polynucleotide of the present invention shall be provided, preferably, either as an isolated polynucleotide (i.e. isolated from its natural context) or in genetically modified form. The polynucleotide, preferably, is DNA, including cDNA, or is RNA. The term encompasses single as well as double stranded polynucleotides. Preferably, the polynucleotide is a chimeric molecule, i.e., preferably, comprises at least one nucleic acid sequence, preferably of at least 20 bp, more preferably at least 100 bp, heterologous to the residual nucleic
30

acid sequences. Moreover, preferably, comprised are also chemically modified polynucleotides including naturally occurring modified polynucleotides such as glycosylated or methylated polynucleotides or artificial modified one such as biotinylated polynucleotides.

- 5 The present invention also relates to a method of identifying at least one biomarker of reactive T-cells, comprising
- (I) providing expression data of a plurality of biomarkers of T-cells in a sample of a subject,
 - (II) providing a clustering said plurality of T-cells based on the expression of the biomarkers of step (A);
 - 10 (III) providing amino acid sequences of at least the complementarity determining regions (CDRs) of TCR chains of T-cells of step (B);
 - (IV) determining recognition of cancer cells by a TCR comprising the complementarity determining regions (CDRs) of step (C);
 - (V) repeating steps (C) and (D) at least once for further T-cells clustering with T-cells whose
 - 15 TCRs are determined to recognize cells presenting a T-cell activating antigen in step (D), wherein the TCRs of said further T-cells are non-identical to the TCRs of step (D);
 - (VI) determining at least one cluster of step (B) comprising the highest fraction of T-cells comprising T-cell receptors recognizing cells presenting a T-cell activating antigen; and
 - (VII) determining at least one biomarker expressed by the highest fraction of T-cells in the
 - 20 cluster determined in step (F), thereby identifying at least one biomarkers of reactive T-cells.

The method of the present invention, preferably, is an in vitro method. Moreover, it may comprise steps in addition to those explicitly mentioned above. Moreover, one or more of said steps may be aided or performed by automated equipment.

25

- The terms "providing expression data" and "providing amino acid sequences" are understood by the skilled person to include each and every means of making the respective data available. Such data may be provided from pre-existing databases, preferably expression databases. Preferably, providing expression data of a plurality of biomarkers of T-cells comprises determining
- 30 expression of said biomarkers e.g. by hybridization of RNA or cDNA derived therefrom to an expression array according to methods known in the art. As referred to herein, providing expression data is providing expression data for single cells, i.e. providing expression data of the biomarkers for each cell separately, thus allowing identification sets of biomarkers expressed

by a T-cell. Thus, expression data are preferably determined by single-cell determination of gene expression, more preferably by single-cell RNA sequencing, as specified elsewhere herein. Preferably, the expression data comprise the sequences of at least the CDRs of the TCRs expressed by said T-cells. Preferably, the expression data comprise expression data of T-cell
5 activation biomarkers and/or of the biomarkers specified herein above.

The term "providing a clustering" relates to providing an allocation of individual T-cells into clusters sharing similar sets of expressed biomarkers. Said clustering preferably is performed in a computer-implemented manner by an algorithm known in the art, e.g. graph-based cluster-
10 ing or k-mean clustering MacQueen (1967), "Some methods for classification and analysis of multivariate observations", 5th Berkeley Symposium on Mathematical Statistics and Probability. Clustering may be visualized by methods also known in the art, e.g. tSNE (van der Maaten and Hinton (2008), J Machine Learning Res 9:2579) or UMAP (McInnes et al. (2020), arXiv:1802.03426v3), preferably UMAP. However, other clustering methods may be used as
15 well. Preferably, a multitude of clusters, i.e. at least two, preferably at least five, more preferably at least ten, still more preferably at least 25, is provided. Preferably, the clustering, i.e. the result of the clustering step, is subject-specific.

The term "determining at least one cluster" is understood by the skilled person. According to
20 step (II) of the method, a clustering is provided, preferably providing at least two clusters; according to step (IV), members of the at least two clusters are evaluated whether they are reactive T-cells, and according to step (VI), at least one cluster is determined comprising the highest fraction of T-cells comprising T-cell receptors recognizing cells presenting a T-cell activating antigen. As the skilled person understands, this step preferably identifies the cluster(s) compris-
25 ing reactive T-cells without the need to know initially which biomarkers are indicative of reactive T-cells. Once a cluster is identified, further T-cell members of the same cluster will preferably be assumed to also be cancer reactive. As will also be understood, repeating steps (III) and (IV) at least once, preferably at least twice, more preferably at least three times, allows for further refining the cluster definition. The biomarkers expressed with the highest frequency in
30 the cluster(s) eventually identified will preferably be assumed to be biomarkers of cancer - reactive T-cells.

The invention further discloses and proposes a computer program including computer-executable instructions for performing the method according to the present invention in one or more of the embodiments enclosed herein when the program is executed on a computer or computer network. Specifically, the computer program may be stored on a computer-readable data carrier.

5 Thus, specifically, one, more than one or even all of method steps a) to d) as indicated above may be performed by using a computer or a computer network, preferably by using a computer program.

The invention further discloses and proposes a computer program product having program code means, in order to perform the method according to the present invention in one or more of the embodiments enclosed herein when the program is executed on a computer or computer network. Specifically, the program code means may be stored on a computer-readable data carrier. Further, the invention discloses and proposes a data carrier having a data structure stored thereon, which, after loading into a computer or computer network, such as into a working
10 memory or main memory of the computer or computer network, may execute the method according to one or more of the embodiments disclosed herein.
15

The invention further proposes and discloses a computer program product with program code means stored on a machine-readable carrier, in order to perform the method according to one or more of the embodiments disclosed herein, when the program is executed on a computer or computer network. As used herein, a computer program product refers to the program as a tradable product. The product may generally exist in an arbitrary format, such as in a paper format, or on a computer-readable data carrier. Specifically, the computer program product may be distributed over a data network.

25

Finally, the invention proposes and discloses a modulated data signal which contains instructions readable by a computer system or computer network, for performing the method according to one or more of the embodiments disclosed herein.

30 Preferably, referring to the computer-implemented aspects of the invention, one or more of the method steps or even all of the method steps of the method according to one or more of the embodiments disclosed herein may be performed by using a computer or computer network. Thus, generally, any of the method steps including provision and/or manipulation of data may

be performed by using a computer or computer network. Generally, these method steps may include any of the method steps, typically except for method steps requiring manual work, such as providing the samples and/or certain aspects of performing the actual measurements.

5 Specifically, the present invention further discloses:

- A computer or computer network comprising at least one processor, wherein the processor is adapted to perform the method according to one of the embodiments described in this description,
- 10 - a computer loadable data structure that is adapted to perform the method according to one of the embodiments described in this description while the data structure is being executed on a computer,
- a computer program, wherein the computer program is adapted to perform the method according to one of the embodiments described in this description while the program is being executed on a computer,
- 15 - a computer program comprising program means for performing the method according to one of the embodiments described in this description while the computer program is being executed on a computer or on a computer network,
- a computer program comprising program means according to the preceding embodiment, wherein the program means are stored on a storage medium readable to a computer,
- 20 - a storage medium, wherein a data structure is stored on the storage medium and wherein the data structure is adapted to perform the method according to one of the embodiments described in this description after having been loaded into a main and/or working storage of a computer or of a computer network, and
- 25 - a computer program product having program code means, wherein the program code means can be stored or are stored on a storage medium, for performing the method according to one of the embodiments described in this description, if the program code means are executed on a computer or on a computer network.

30 In view of the above, the following embodiments are particularly envisaged:

Embodiment 1: A method of identifying a T-cell reactive to cells of a subject presenting a T-cell activating antigen (reactive T-cell), comprising

(a) determining expression of at least one of CCL4, CCL4L2, CCL3, CCL3L1, and CXCL13 in T-cells from a sample of said subject; and

(b) identifying a reactive T-cell based on the determination of step (a), preferably wherein said T-cell activating antigen is a cancer antigen or an autoimmune T-cell antigen, more preferably is a cancer antigen.

Embodiment 2: A method of identifying a T-cell reactive to cancer cells (cancer-reactive T-cell), comprising

(a) determining expression of at least one of CCL4, CCL4L2, CCL3, CCL3L1, and CXCL13 in T-cells from a sample of a subject; and

(b) identifying a cancer-reactive T-cell based on the determination of step (a).

Embodiment 2: The method of embodiment 1 or 2, wherein step (a) comprises determining expression of at least two, preferably at least three, more preferably at least four, of CCL4, CCL4L2, CCL3, CCL3L1, and CXCL13.

Embodiment 4: The method of any one of embodiments 1 to 3, wherein step (a) comprises

further determining expression of at least one biomarker selected from the list consisting of IFNG, HAVCR2, FNBP1, CSRNPI, SPRY1, RHOH, FOXN2, HIF1A, TOB1, RILPL2, CD8B, GABARAPL1, TNFSF14, EGR1, EGR2, TAGAP, TNFSF9, ANXA1, MAP3K8, PIK3R1, DUSP2, DUSP4, DUSP6, CLIC3, RASGEF1B, LAG3, XCL2, NR4A2, DNAJB6, NFKBID, MCL1, EVI2A, SLC7A5, H3F3B, NR4A3, REL, IRF4, CST7, ATF3, TNF, GPR171, BCL2A1, ITGA1, TNFAIP3, NR4A1, RUNX3, HERPUD2, FASLG, CBLB, PTGER4, SLA, XCL1, BHLHE40, LYST, KLRD1, ZNF682, CTSW, SLC2A3, NLRP3, SCML4, VSIR, LINC01871, and ZFP36L1.

Embodiment 5: The method of any one of embodiments 1 to 4, wherein expression is determined in step (a) by single-cell determination of gene expression, preferably by single-cell RNA sequencing and/or wherein said sample is a tumor sample.

Embodiment 6: A method of identifying a TCR binding to an activating antigen presented on a cell, preferably a cancer cell, of a subject, said method comprising

(A) identifying a reactive T-cell according to the method of any one of embodiments 1 to 5,

(B) providing the amino acid sequences of at least the complementarity determining regions

(CDRs) of the TCR of the reactive T-cell identified in step (A); and, hereby,

(C) identifying a TCR binding to an activating antigen presented on a cell.

Embodiment 7: The method of any one of embodiment 1 to 6, wherein at least one biomarker of step a) and/or the nucleic acid sequences in step (B) is/are determined by single-cell sequencing, preferably by single-cell RNA sequencing.

Embodiment 8: The method of embodiment 6 or 7, wherein said method comprises further step
5 B1) expressing a TCR comprising at least the CDRs determined in step B) in a host cell, preferably a T-cell.

Embodiment 9: The method of any one of embodiments 6 to 8, wherein said method comprises further step B1) expressing a TCR comprising at least the CDRs determined in step B) in a host cell, preferably a T-cell.

10 Embodiment 10: The method of embodiment 9, wherein said method comprises further step B2) determining binding of the TCR expressed in step B1) to an activating antigen presented on a cell, preferably a cancer antigen, complexed in a major histocompatibility complex (MHC), preferably MHC class I, molecule, preferably in a tetramer assay.

Embodiment 11: The method of embodiment 9 or 10, wherein said method further comprises
15 step B3) determining recognition of cells presenting a T-cell activating antigen by the TCR expressed in step B1).

Embodiment 12: The method of any one of embodiments 6 to 11, wherein said method further comprises further step B4) producing a soluble TCR comprising at least the CDRs determined in step B) and determining binding of said soluble TCR to a T-cell activating antigen; preferably
20 to a cancer antigen complexed in a major histocompatibility complex (MHC), preferably MHC class I, molecule.

Embodiment 13: A method of providing a T-cell recognizing a cell presenting a T-cell activating antigen, preferably a cancer cell, said method comprising

(i) identifying a TCR binding to a cell presenting a T-cell activating antigen according to the
25 method according to any one of embodiments 6 to 12,
(ii) expressing a TCR comprising at least the complementarity determining regions (CDRs) of the TCR of step (I) in a T-cell, and, thereby,
(iii) providing a T-cell recognizing a cell presenting a T-cell activating antigen, preferably a cancer cell.

30 Embodiment 14: The method of embodiment 13, wherein said method further comprises a step of testing reactivity of the T-cell of step (ii) to cells presenting a T-cell activating antigen.

Embodiment 15: The method of any one of embodiments 1 to 14, wherein said sample is a tissue sample or a bodily fluid sample.

- Embodiment 16: The method of any one of embodiments 1 to 15, wherein said sample is a blood sample.
- Embodiment 17: The method of any one of embodiments 1 to 16, wherein said sample is a cancer sample.
- 5 Embodiment 18: The method of any one of embodiments 1 to 17, wherein said sample is a sample of non-cancer tissue, preferably of cancer-adjacent tissue.
- Embodiment 19: A reactive T-cell identified by the method according to any one of embodiments 1 to 5 and/or obtained or obtainable by the method according to embodiment 13 or 14, preferably comprising a T-cell receptor comprising an amino acid sequence of SEQ ID NO:1
- 10 and/or SEQ ID NO:2, for use in medicine.
- Embodiment 20: A reactive T-cell identified by the method according to any one of embodiments 1 to 5 and/or obtained or obtainable by the method according to embodiment 13 or 14, preferably comprising a T-cell receptor comprising an amino acid sequence of SEQ ID NO:1 and/or SEQ ID NO:2, for use in treating and/or preventing cancer in a subject.
- 15 Embodiment 21: The subject matter of any one of embodiments 1 to 20, wherein said subject is an apparently healthy subject.
- Embodiment 22: The subject matter of any one of embodiments 1 to 21, wherein said subject is a subject afflicted with cancer.
- Embodiment 23: The subject matter of any one of embodiments 1 to 22, wherein said cells
- 20 presenting a T-cell activating antigen are cancer cells, preferably tumor cells.
- Embodiment 24: A method of identifying at least one biomarker of reactive T-cells, comprising
- (I) providing expression data of a plurality of biomarkers of T-cells in a sample of a subject,
- (II) providing a clustering said plurality of T-cells based on the expression of the biomarkers of step (I);
- 25 (III) providing amino acid sequences of at least the complementarity determining regions (CDRs) of TCRs of T-cells of step (II);
- (IV) determining reactivity of T-cells expressing a TCR comprising the CDRs of step (III) to cells presenting a T-cell activating antigen;
- (V) repeating steps (III) and (IV) at least once for further T-cells clustering with T-cells whose
- 30 TCRs are determined to be reactive to cells presenting a T-cell activating antigen in step (IV), wherein the TCRs of said further T-cells are non-identical to the TCRs of step (IV);
- (VI) determining at least one cluster of step (II) comprising the highest fraction of T-cells comprising T-cell receptors recognizing cells presenting a T-cell activating antigen; and

(VII) determining at least one biomarker expressed by the highest fraction of T-cells in the cluster determined in step (VI), thereby identifying at least one biomarkers of cancer-reactive T-cells.

Embodiment 25: The subject matter of any of the preceding embodiments, wherein said T-cell(s) is/are are CD8⁺ T-cell(s) or CD4⁺ T-cells, preferably are CD8⁺ T-cells.

Embodiment 26: The subject matter of any of the preceding embodiments, wherein said TCR comprises, preferably consists of, a TCR alpha chain and a TCR beta chain or a TCR gamma chain and a TCR delta chain, preferably comprises, more preferably consists of, a TCR alpha chain and a TCR beta chain.

Embodiment 27: A polynucleotide encoding at least one TCR TCR binding to an activating antigen provided or identifiable according to the method according to any one of embodiments 6 to 12.

Embodiment 28: The subject matter of any of the preceding embodiments, wherein said reactive T-cell is a cancer-reactive T-cell.

Embodiment 29: The method of any one of embodiments 1 to 3, wherein step (a) comprises further determining expression of at least one biomarker selected from the list consisting of LAG3, GABARAPL1, CBLB, SLA, KLRD1, and CLEC2B, preferably comprises determining all biomarkers of embodiment 1 and/or of embodiment 29.

Embodiment 30: The method of any one of embodiments 1 to 3 and 29, wherein step (a) comprises further determining expression of at least one biomarker selected from the list consisting of CTSD, CD7, CD3D, LSP1, SNAP47, GAPDH, KLRK1, TNS3, VCAM1, KLRC2, PMAIP1, FYN, CTLA4, GSTP1, AREG, FAM3C, SH3BGRL3, CD3E, SRGAP3, SRGN, SIRPG, SCPEP1, RHOB, ANKRD28, LINC02446, RABAC1, IKZF3, BCAS4, CD2, BLOC1S1, RHOA, EID1, MYL6, CLIC1, IQGAP1, ARPC2, PHYKPL, PRDM1, EVL, TPI1, ADGRE5, PAXX, RGS2, HERPUD1, IFI27L2, SEPTIN7, UBB, JUN, CFLAR, LITAF, ANXA5, STAT3, RSRP1, PRDX5, SEM1, SER-PINB1, RNF19A, IL2RG, ENSA, SRP14, ATP6V0C, LY6E, BIN1, AKAP13, PDE4D, PELI1, PARK7, MSN, SERTAD1, RAC2, SELENOH, PSMB8, CKLF, KLRC1, RNASEK, MT2A, TXNIP, and FOXP3.

Embodiment 31: The method of any one of embodiments 1 to 3 and 29 to 30, wherein step (a) comprises determining expression of at least CCL3L1 + CCL3; CCL3L1 + CCL3 + LAG3 + KLRD1; CCL3L1 + CCL3 + CXCL13; CCL3L1 + CCL3 + CXCL13 + KLRD1; CCL3L1 + CCL3 + CXCL13 + LAG3; CCL3L1 + CCL3 + CXCL13 + LAG3 + KLRD1; CCL3L1 + CCL3 + KLRD1; CCL3L1 + CCL3 + LAG3; CCL3L1 + CCL4 + CCL3; CCL3L1 + CCL4 +

CCL3 + CXCL13; CCL3L1 + CCL4 + CCL3 + CXCL13 + KLRD1; CCL3L1 + CCL4 +
 CCL3 + CXCL13 + LAG3; CCL3L1 + CCL4 + CCL3 + CXCL13 + LAG3 + KLRD1;
 CCL3L1 + CCL4 + CCL3 + LAG3; CCL3L1 + CCL4 + CCL3 + LAG3 + KLRD1; CCL3L1
 + CCL4 + CCL4L2 + CCL3 + CXCL13 + KLRD1; CCL3L1 + CCL4 + CCL4L2 + CCL3 +
 5 CXCL13 + LAG3 ; CCL3L1 + CCL4 + CCL4L2 + CCL3 + CXCL13 + LAG3 + KLRD1;
 CCL3L1 + CCL4 + CCL4L2 + CCL3 + KLRD1; CCL3L1 + CCL4 + CCL4L2 + CXCL13;
 CCL3L1 + CCL4 + CCL4L2 + CXCL13 + KLRD1; CCL3L1 + CCL4 + CCL4L2 + CXCL13
 + LAG3; CCL3L1 + CCL4 + CCL4L2 + CXCL13 + LAG3 + KLRD1; CCL3L1 + CCL4 +
 CCL4L2 + LAG3 + KLRD1; CCL3L1 + CCL4 + CXCL13; CCL3L1 + CCL4 + CXCL13 +
 10 KLRD1; CCL3L1 + CCL4 + CXCL13 + LAG3; CCL3L1 + CCL4 + CXCL13 + LAG3 +
 KLRD1; CCL3L1 + CCL4 + KLRD1; CCL3L1 + CCL4 + LAG3 + KLRD1; CCL3L1 +
 CCL4L2 + CCL3; CCL3L1 + CCL4L2 + CCL3 + CXCL13; CCL3L1 + CCL4L2 + CCL3 +
 CXCL13 + KLRD1; CCL3L1 + CCL4L2 + CCL3 + CXCL13 + LAG3; CCL3L1 + CCL4L2
 + CCL3 + CXCL13 + LAG3 + KLRD1; CCL3L1 + CCL4L2 + CCL3 + KLRD1; CCL3L1 +
 15 CCL4L2 + CCL3 + LAG3 + KLRD1; CCL3L1 + CCL4L2 + CXCL13; CCL3L1 + CCL4L2
 + CXCL13 + KLRD1; CCL3L1 + CCL4L2 + CXCL13 + LAG3; CCL3L1 + CCL4L2 +
 CXCL13 + LAG3 + KLRD1; CCL3L1 + CCL4L2 + KLRD1; CCL3L1 + CCL4L2 + LAG3;
 CCL3L1 + CCL4L2 + LAG3 + KLRD1; CCL3L1 + CXCL13; CCL3L1 + CXCL13 +
 KLRD1; CCL3L1 + CXCL13 + LAG3; CCL3L1 + CXCL13 + LAG3 + KLRD1; CCL3L1 +
 20 LAG3 + KLRD1; CCL3 + CXCL13; CCL3 + CXCL13 + KLRD1; CCL3 + CXCL13 +
 LAG3; CCL3 + CXCL13 + LAG3 + KLRD1; CCL3 + LAG3 + KLRD1; CCL4 + CCL3;
 CCL4 + CCL3 + CXCL13; CCL4 + CCL3 + CXCL13 + KLRD1; CCL4 + CCL3 + CXCL13
 + LAG3; CCL4 + CCL3 + CXCL13 + LAG3 + KLRD1; CCL4 + CCL3 + KLRD1; CCL4 +
 CCL3 + LAG3 + KLRD1; CCL4 + CCL4L2 + CCL3 + CXCL13; CCL4 + CCL4L2 + CCL3
 25 + CXCL13 + KLRD1; CCL4 + CCL4L2 + CCL3 + CXCL13 + LAG3; CCL4 + CCL4L2 +
 CCL3 + CXCL13 + LAG3 + KLRD1; CCL4 + CCL4L2 + CCL3 + KLRD1; CCL4 +
 CCL4L2 + CCL3 + LAG3; CCL4 + CCL4L2 + CXCL13; CCL4 + CCL4L2 + CXCL13 +
 KLRD1; CCL4 + CCL4L2 + CXCL13 + LAG3; CCL4 + CCL4L2 + CXCL13 + LAG3 +
 KLRD1; CCL4 + CCL4L2 + KLRD1; CCL4 + CCL4L2 + LAG3 + KLRD1; CCL4 +
 30 CXCL13; CCL4 + CXCL13 + KLRD1; CCL4 + CXCL13 + LAG3; CCL4 + CXCL13 +
 LAG3 + KLRD1; CCL4 + LAG3 + KLRD1; CCL4L2 + CCL3; CCL4L2 + CCL3 +
 CXCL13; CCL4L2 + CCL3 + CXCL13 + KLRD1; CCL4L2 + CCL3 + CXCL13 + LAG3;
 CCL4L2 + CCL3 + CXCL13 + LAG3 + KLRD1; CCL4L2 + CCL3 + KLRD1; CCL4L2 +

CCL3 + LAG3 + KLRD1; CCL4L2 + CXCL13; CCL4L2 + CXCL13 + KLRD1; CCL4L2 + CXCL13 + LAG3; CCL4L2 + CXCL13 + LAG3 + KLRD1; CCL4L2 + LAG3 + KLRD1; CXCL13 + LAG3; CXCL13 + LAG3 + KLRD1; KLRD1; KLRD1 + CCL3; KLRD1 + CCL3L1; KLRD1 + CCL4L2; KLRD1 + CXCL13; KLRD1 + LAG3; all biomarkers of Table 1; all biomarkers of Table 5; or all biomarkers of Table 6.

Embodiment 32: The method of any one of embodiments 1 to 3 and 29 to 31, wherein said T-cell activating antigen is a cancer antigen, and wherein preferably said sample is a tumor sample.

Embodiment 33: The method of embodiment 32, wherein said cancer is a brain metastasis of a non-brain primary tumor, is lung cancer, or is glioblastoma, preferably is a brain metastasis of a non-brain primary tumor or is lung cancer.

Embodiment 34: A method of identifying a TCR binding to a T-cell activating antigen presented on a cell, preferably a cancer cell, of a subject, said method comprising

(A) identifying a reactive T-cell according to the method of any one of embodiments 1 to 3 and 29 to 33,

(B) providing the amino acid sequences of at least the complementarity determining regions (CDRs) of the TCR of the reactive T-cell identified in step (A); and, hereby,

(C) identifying a TCR binding to an activating antigen presented on a cell.

Embodiment 35: The method of any one of embodiments 1 to 3 and 29 to 34, wherein expression of at least one biomarker of step a) and/or the nucleic acid sequences encoding the amino acid sequences of step (B) is/are determined by single-cell sequencing, preferably by single-cell RNA sequencing.

Embodiment 36: The method of embodiment 34 or 35, wherein said method comprises further step B1) expressing a TCR comprising at least the CDRs determined in step B) in a host cell, preferably a T-cell.

Embodiment 37: The method of embodiment 36, wherein said method further comprises further step B2) determining binding of the TCR expressed in step B1) to a T-cell activating antigen, preferably complexed in a major histocompatibility complex (MHC), preferably MHC class I, molecule, preferably in a tetramer assay.

Embodiment 38: The method of embodiment 36 or 37, wherein said method further comprises step B3) determining recognition of cells presenting a T-cell activating antigen by the TCR expressed in step B1).

Embodiment 39: The method of any one of embodiments 34 to 38, wherein said method further comprises step B4) producing a soluble TCR comprising at least the CDRs determined in step B) and determining binding of said soluble TCR to a cancer cell and/or to a cancer antigen complexed in a major histocompatibility complex (MHC), preferably MHC class I, molecule.

5 Embodiment 40: A method of providing a T-cell recognizing a cell presenting a T-cell activating antigen, preferably a cancer cell, said method comprising

(i) identifying a TCR binding to a cell presenting a T-cell activating antigen according to the method according to any one of embodiments 34 to 36,

(ii) expressing a TCR comprising at least the complementarity determining regions (CDRs) of the TCR of step (I) in a T-cell, and, thereby,

10 (iii) providing a T-cell recognizing a cell presenting a T-cell activating antigen, preferably a cancer cell.

Embodiment 41: A reactive T-cell identified by the method according to any one of embodiments 1 to 3 and 29 to 33 and/or obtained or obtainable by the method according to any one of
15 embodiments 34 to 40, preferably comprising a T-cell receptor comprising an amino acid sequence of SEQ ID NO:1 and/or SEQ ID NO:2, for use in medicine or for use in treating and/or preventing cancer in a subject.

Embodiment 42: A method of identifying at least one biomarker of reactive T-cells, comprising (I) providing expression data of a plurality of biomarkers of T-cells in a sample of a subject,

20 (II) providing a clustering said plurality of T-cells based on the expression of the biomarkers of step (I);

(III) providing amino acid sequences of at least the complementarity determining regions (CDRs) of TCRs of T-cells of step (II);

(IV) determining reactivity of T-cells expressing a TCR comprising the CDRs of step (III) to
25 cells presenting a T-cell activating antigen;

(V) repeating steps (III) and (IV) at least once for further T-cells clustering with T-cells whose TCRs are determined to be reactive to cells presenting a T-cell activating antigen in step (IV), wherein the TCRs of said further T-cells are non-identical to the TCRs of step (IV);

(VI) determining at least one cluster of step (II) comprising the highest fraction of T-cells comprising T-cell receptors recognizing cells presenting a T-cell activating antigen; and
30

(VII) determining at least one biomarker expressed by the highest fraction of T-cells in the cluster determined in step (VI), thereby identifying at least one biomarkers of cancer-reactive T-cells.

Embodiment 43: The subject matter of any one of embodiments 1 to 3 and 29 to 42, wherein said T-cell(s) is/are CD8⁺ T-cell(s) or CD4⁺ T-cells, preferably are CD8⁺ T-cell(s).

Embodiment 44: The subject matter of any one of embodiments 34 to 43, wherein said TCR comprises, preferably consists of, a TCR alpha chain and a TCR beta chain.

- 5 Embodiment 44: A method of identifying a T-cell reactive to cells of a subject presenting a T-cell activating antigen (reactive T-cell), comprising
- (a) determining expression of at least one of KLRD1 and LAG3 in T-cells from a sample of said subject; and
 - (b) identifying a reactive T-cell based on the determination of step (a),
- 10 preferably wherein said T-cell activating antigen is a cancer antigen or an autoimmune T-cell antigen, more preferably is a cancer antigen.

Embodiment 45: The subject matter of embodiment 44, further comprising at least one feature of any one of embodiments 1 to 43.

15

All references cited in this specification are herewith incorporated by reference with respect to their entire disclosure content and the disclosure content specifically mentioned in this specification.

20 Figure Legends

Figure 1: A) and B) show results of UMAP clustering of T Cells for 2 patients separately. D)-F) and Figure G)- J) show the expression of the core genes CCL3, CCL3L1, CCL4 and CCL4L2 in the clustered cells, respectively, for patient 1 (D)- F)) and Patient 2 (G)- J)), (K) shows the expression of the core gene CXCL13 in Patient 2.

25

Figure 2: shows the clusters of cancer-reactive T-cells defined based on the expression of core genes CCL3, CCL3L1, CCL4 and CCL4L2 for Patient 1 (A)) and Patient 2 (B) whereas C) shows the cluster of cancer-reactive T-cells defined based on the expression core gene CXCL13 in Patient 2.

30

Figure 3: A) shows the distribution of selected TCR clones (X-axis) in transcriptomic clusters (Y-axis) for Patient 1. B) shows the clustering of TCR based on the TCR fraction in the reactive cluster and C) shows the TCR testing result based on FACS based assay.

5 Figure 4: A) shows the distribution of selected TCR clones (X-axis) in transcriptomic clusters (Y-axis) for Patient 2. B) shows the clustering of TCR based on the TCR fraction in the reactive cluster and C) shows the TCR testing result based on NFAT reporter assay: Co-culture of TCR transgenic Jurkat cells with peptide-loaded autologous PBMCs confirms that TCR4 recognizes the IDH1.R132H mutant epitope expressed by the tumor. Data depicted as mean + SD of 3
10 technical replicates. Representative of 3 independent experiments. CD3+CD28 stimulation represents the maximum possible activation of T cells. MOG is negative control peptide not bound by either TCR in the assay.

Figure 5: A) Overview of the process; Patient T cells are isolated from tumor material and their
15 transcriptome and VDJ sequences determined; using a classifier, T cells with a "reactive signature" are identified, optionally using a multitude of markers on a UMAP; b) Figure 1B: Example of in vitro testing results.

Figure 6: A) Testing data of Patient 3 (non-primary brain metastasis); black dots show reactive
20 T-cells, bright gray dots show untested T-cells, and dark gray dots show T-cells tested non-reactive; the solid line polygon shows the area in which tested reactive T-cells cluster, the dotted line polygon shows a comparison area in which tested non-reactive T-cells cluster; B)-E) Prediction of Reactive TCR using 5 core genes (B), 7 alternative core genes (C), core and accessory genes (D), genes of signature 2 (E); F) and G) are comparative Examples using the biomarkers
25 identified in Lowery et (2022)(F) and WO 2021/188954 A1 (G), respectively; H) - L): expression of core genes H) CCL3L1, I) CCL3, J) CCL4, K) CCL4L2, and L) CXCL13 for Patient 3.

Figure 7: A) Testing data of Patient 2 (glioma); black dots show reactive T-cells, bright gray
30 dots show untested T-cells, and dark gray dots show T-cells tested non-reactive; B)-E) Prediction of Reactive TCR using 5 core genes (B), 7 alternative core genes (C), core and accessory genes (D), and genes of signature 2 (E); F) and G) are comparative Examples using the biomarkers identified in Lowery et (2022)(F) and WO 2021/188954 A1 (G), respectively.

Figure 8: A) External data on T-cells reactive to lung cancer; black dots show reactive T-cells, bright gray dots show untested T-cells; B)-E) and H) Prediction of Reactive TCR using 5 core genes (B), 7 alternative core genes (C), core and accessory genes (D), genes of signature 2 ((E), and CCL3L1 alone (H); F) and G) are comparative Examples using the biomarkers identified in Lowery et (2022)(F) and WO 2021/188954 A1 (G), respectively.

The following Examples shall merely illustrate the invention. They shall not be construed, whatsoever, to limit the scope of the invention.

10 **Example 1: Single Cell Library Preparation**

Single cell suspension of tumor was FACS-sorted for CD45⁺CD3⁺ population to enrich for T Cells. Single cell library construction of sorted T Cells was performed using Chromium Single Cell Immune Profiling Kit (10X Chromium) according to the manufacturer's protocol. The constructed scVDJ and scRNA library were then sequenced on Hiseq2500 Rapid / Nextseq550 and Hiseq4000 (Illumina) respectively.

Example 2: Single Cell RNA Analysis

Sequencing Raw data was processed with cellranger pipeline (v3.1.0) with corresponding GRCh38 genome assembly with default settings to generate gene expression matrices. Matrices were imported into R and analyzed using the Seurat package. For quality control, outliers were removed based on UMI, the number of genes and the percentage of mitochondrial gene expression. Then, gene expression was transformed and normalized and VDJ genes were then subsequently removed from the variable genes. Highly variable genes were selected based on Principal Component Analysis and the number of components was selected based on inflection point in the elbow plot. Cells were then clustered using unsupervised graph-based clustering method and UMAP were plotted for visualization. Differential gene expression analysis was done using MAST and the upregulated genes were used to define each cluster.

The scVDJ data was processed similarly using the cellranger pipeline with default settings. The T cell receptor data was then mapped onto gene expression data to determine the distribution of individual TCR clones transcriptomically. K-mean clustering was done to cluster the TCR based on their distribution in transcriptomic clusters.

Example 3: Cloning

For cloning the TCRs, synthetic alpha and beta VDJ fragments of the variable region of the TCR were obtained from Twist Biosciences. The TCR variable fragments were inserted into an S/MAR sequence-bearing expression vector (pSMARTer) that allows extrachromosomal replication of the vector in eukaryotic cells using a single-step Bsa-I mediated Golden Gate reaction. The expression vector was designed to harbor murine alpha and beta constant TCR regions and a p2a self-cleaving peptide linker to facilitate production of separate alpha and beta polypeptide chains of the TCR. The vector was subsequently transformed into NEB5-alpha competent E.coli (NEB); colonies screened for transgene by antibiotic resistance; and endotoxin-free plasmid prepared using NucleoBond Extra Maxi EF kit (Macherey-Nagel) for transfection.

Example 4: NFAT Assay

The cloned TCR expression vector and a nano-luciferase-based NFAT reporter vector (pDONR, with 4x NFAT-response elements) were transfected into Jurkat Δ76 cells using electroporation (Neon Transfection system, ThermoFisher Scientific). In brief, 2×10^6 cells were used per electroporation with Neon 100 μl tips (8 μg TCR expression vector + 5 μg NFAT reporter vector). Cells were harvested and washed based on manufacturer's protocol, then electroporated with 1325V, 10ms, 3 pulses and transferred to antibiotic-free RPMI 1640 medium containing 10% FCS. Patient-autologous PBMCs were used as antigen presenting cells (APCs) and thawed 24 h before co-culture in X-VIVO 15 medium (Lonza) containing 50U/ml Benzonase (Sigma-Aldrich), and rested for 6-8 h before seeding into 96-well white-opaque tissue culture treated plates (Falcon) at 1.5×10^5 cells per well. Cells were loaded with peptides at a final concentration of 10 μg/ml in a total volume of 150 μl for 16h. Peptides utilized were human IDH1R132H peptides (p123-142), MOG (p35-55) at equal concentrations and PBS + 10% DMSO (vehicle) at equal volume as negative controls. 48h post electroporation, TCR-transgenic Jurkat Δ76 cells were harvested and co-cultured with peptide-loaded PBMC for 6h at a 1:1 ratio. Human T-cell TransAct beads (Miltenyi) were used as positive control. A publicly known TCR against InfluenzaHA (p307-319) was used as an assay reference. Nano-luciferase induction indicating TCR activation was assayed using Nano-Glo Luciferase assay system (Promega) according to manufacturer's protocol and signal was detected on PHERAstar FS plate reader (BMG Labtech).

Example 5: FACS-based Assay

Cloning was done as described above with the addition of T7 promoter by PCR. The PCR product was then purified using DNA Clean & Concentrator-5 (Zymo Research) and used as a template for *in vitro* transcription using Cellscript kit according to the manufacturer's protocol.

- 5 The concentration and integrity of RNA was assessed by Nanodrop and Bioanalyzer respectively. The RNA was then electroporated into expanded autologous PBMC using the Lonza 4D nucleofactor device. After electroporation, cells were incubated at room temperature for 10 minutes before plating into 48-well plate containing 1mL of media (TexMACS + 2% AB) and allowed to rest overnight. Prior to incubation, 150k cells were stained with CD3, CD4, CD8a, mTCR β to use as a control. The rest of the electroporated cells were then co-incubated with target cells (tumour cell line/patient derived xenograft) for 5 hours, with the addition of Golgistop and Golgiplug after 1 hour. After co-incubation, cells were stained for dead cell biomarker, CD3, CD4, CD8a, mTCR β , TNF α , and IFN γ , and then measured using FACSLytic (BD Biosciences). The analysis was done using FlowJo. In the exemplary analysis underlying
- 10 Fig. 6B, cells were stained for dead cell biomarker, CD3, CD4, CD8a, mTCR β , TNF α , and CD107a and then measured using FACSLytic.

Example 6: Results-1

20 **6.1 Clustering of T Cells based on gene expression**

Single cell RNA-seq dataset was normalized, transformed and clustered using graph-based unsupervised clustering. 2 selected patients data are shown here. 15 clusters and 16 clusters were identified from Patient 1 (Figure 1A) and Patient 2 (Figure 1B), respectively.

25 **6.2 Expression of signature genes**

Differential gene expression was done using MAST and the upregulated gene expression for each cluster was found. In multiple patients, we have identified a cluster that expressed the signature genes CCL3, CCL3L1, CCL4 and CCL4L2. Figure 1D-1F and Figure 1G-1J show the expression of the signature genes of 2 selected patients. Another signature gene CXCL13 expression is also shown in a selected patient (Figure 1K).

30

6.3 Defining reactive clusters based on signature genes

Based on the expression of signature genes CCL3, CCL3L1, CCL4 and CCL4L2, a reactive cluster was defined (Figure 2A and 2B). The reactive cluster defined based on the expression of the signature gene CXCL13 was depicted in Figure 2C.

5

6.4 CCL3/CCL3L1/CCL4/CCL4L2 Signature

Figure 3A depicts the top 13 highest frequency TCR clonotypes in Patient 1. As described previously, cluster 4 which expressed the signature genes was defined as the signature cluster. From the distribution, we can clearly see that TCR1, TCR12 and TCR13 have higher distribution of T Cell in signature cluster.

10

Figure 3B shows the k-mean clustering result based on the fraction of T Cells in the signature cluster. 3 clusters were found based on this clustering result. Cluster with high fraction of T Cells in the signature cluster should be reactive, cluster with moderate fraction of T Cells should be likely-reactive and the cluster with lowest fraction of T Cells in the signature cluster should be non-reactive. Thus, TCR1 was predicted to be reactive, and TCR12 and TCR13 was predicted to be likely reactive while the other TCR are non-reactive. TCRs were then cloned to test the tumour reactivity of these TCR and to corroborate the TCR prediction based on signature genes.

15
20

Figure 3C shows the result of FACS-based TCR testing. As predicted by the signature genes, only TCR1, TCR13 and possibly TCR12 secrete IFN γ upon coculture with the corresponding patient's tumour cells, thus showing TCR1 and TCR13 are indeed reactive to cancer cells with TCR12 being possibly reactive.

25

6.5 CXCL13 Signature

Figure 4A illustrates the top 5 highest frequency CD4 TCR in Patient 2. From the distribution, it is clear that TCR4 was the only TCR to have higher distribution in the signature cluster. A further k-mean clustering (Figure 4B) also found that there were 2 clusters based on the T Cell fraction in this signature cluster. The cluster with the higher fraction, which consist of only TCR4, was then predicted to be reactive. This TCR was then cloned and tested with NFAT assay. TCR4 which was predicted to be tumour reactive by gene signature is indeed reactive

30

(Figure 4C) when coculture with peptide-loaded PBMC. The exact sequence for this TCR4 is as shown in SEQ ID NOs:1 and 2.

Example 7: Prediction of Reactive TCR based on Signature

The Seurat object from single cell analysis was converted into 'cell_data_set' object from Monocle package in R. A classifier was then trained using the Garnett package in R using the signature genes. Reactive T cells identified by the classifier was then map unto the UMAP to identify the reactive cluster(s). The F-score (a combination of precision and recall) was then calculated using the caret package in R.

Example 8: Results-2

8.1 Prediction of Reactive TCRs in Brain Metastasis Patient

Using the signature genes we identified previously in Patient 1 and Patient 2, we corroborated the predictive power in a separate patient (Patient 3). Infiltrating CD 8+ T cells (TILs) were extracted from a brain metastasis resection of Patient 3 and processed to generate scRNA and scVDJ libraries for sequencing as previously described (Example 1). The resulting data were processed and visualized in 2 dimensions using a UMAP plot, a type of dimensionality reduction (Figure 6A) in which similar cells cluster more closely together than dissimilar cells. Each point on the plot represents a cell, and each cell's gene expression profile and TCR was known. Corroborating results from Patient 1 and Patient 2, the core genes are co-expressed within a specific region (Figure 6H-L).

A number of TCR clonotypes cloned from the TILs were tested for reactivity against the tumor using FACS-based assays (as described in Example 5 above). This reactivity information can then be overlaid onto the UMAP plot; it is evident that TILs expressing tumour-reactive TCR clonotype are predominating clustered within a region enclosed by a solid line on the UMAP plot (Figure 6A).

We went on to elucidate principal genes most important for predicting whether a given T cell clonotype is tumor reactive, our TCR reactivity "signature" (Example 1). The quality of the signature can be approximated computationally by determining how many cells identified by

the signature fall express a TCR known to be reactive to the tumour (which can be graphically approximated by comparing the location of the known tumor reactive TCRs with the predictions in the UMAP plot).

5 We illustrate how different signatures can be mapped onto the UMAP plot, using reactivity prediction based on the 5 core gene signature (Figure 6B), the 7 alternative core gene signature (Figure 6C), the core and accessory gene signature (Figure 6D) and the signature 2 gene signature (Figure 6E). To show the robustness of signature genes prediction, predicted non-reactive TCR clonotypes were also cloned and confirmed to represent *bona fide* non-reactive TCR
10 clonotypes.

Lowery *et. al.* have reported two signatures for reactive T cells, each specific to CD4 or CD8 expressing T cells (Science (2022, comparative Examples A) and WO 2021/188954 A1 (comparative Examples B)). The performance of these signatures was benchmarked against the signature disclosed herein (Table 7) and the new signatures disclosed herein were shown to perform significantly better (i.e. having a higher F-score, i.e. a higher precision and recall in predicting cells expressing a tumour reactive TCR).
15

8.2 Prediction of Reactive TCRs in Glioma Patient

20

Infiltrating T cells (TILs) were extracted from a pseudoprogression sample from a primary glioma patient with an IDH1.R132H mutant tumor, and single cell libraries were prepared and tested as previously described. These TCRs were derived from CD4+ T cells.

25 We found that the gene signature developed by Lowery *et. al.* specifically for CD4+ cells (Figure 7F, 7G) did not perform better than signature 2 gene signature in predicting tumour reactive T cells. Additionally, we found that by limiting our analysis to cells not expressing CD8B, our gene signatures performed much better than the signature reported by Lowery *et. al.* (Figure 7B)-E), Table 9). This shows the general applicability of our gene signature for tumour reactivity across different tumor modalities and for both CD4+ and CD8+ cells.
30

8.3 Prediction of Reactive TCRs in Lung Cancer Patient

The signature of anti tumor reactivity was further validated using the lung cancer dataset published by Caushi *et. al.* (2021), experimentally confirmed tumor reactive TCRs; we mapped this data onto the corresponding UMAP to show that TILs expressing tumour reactive clonotypes (Figure 8A). As previously shown in brain metastasis, our signatures – despite being trained on brain cancer samples - performed consistently better than the signature developed Lowery *et. al.* (Figure 8B-8G, Table 8).

9. Gene lists (signatures) used in the Examples

9.1 "5 Core Genes" ("core"):

CCL3L1, CCL4, CCL4L2, CCL3, and CXCL13.

9.2 "7 Alternative Core Genes" ("core-2"):

CCL3L1, LAG3, GABARAPL1, CBLB, SLA, KLRD1, and CLEC2B.

9.3 Other signatures of the invention:

cf. all biomarkers of Tables 1 to 10 below, respectively.

20

9.4 Comparative signatures:

Genes used in Lowery *et. al.* (2022) (CD8):

ATP10D, GZMB, ENTPD1, KIR2DL4, LAYN, HTRA1, CD70, CXCR6, HMOX1, ADGRG1,
 25 LRRN3, ACP5, CTSW, GALNT2, LINC01480, CARS, LAG3, TOX, PTPRCAP, ASB2,
 ITGB7, PTMS, CD8A, GPR68, NSMCE1, ABI3, SLC1A4, PLEKHF1, CD8B, LINC01871,
 CCL4, NKG7, CLIC3, NDFIP2, PLPP1, PCED1B, CXCL13, PDCD1, PRF1, HLA-DMA,
 GPR25, CD9, TIGIT, HLA-DRB5, SYTL3, SLF1, NEK1, CASP1, SMC4, TSEN54, PLSCR1,
 GNPTAB, HLA-DPB1, PLEKHA1, ARHGAP9, ALOX5AP, SH3BP1, NCF4, NELL2,
 30 GATA3, PPM1M, TNFRSF1A, AC022706.1, MCM5, HLA-DRB1, TNFSF10, TRIM21,
 HDLBP, ERN1, CALHM2, SASH3, ACTA2, MAST4, CAPG, MPST, IGFLR1, GZMA,
 CD27, ITGAE, SLA2, RHOC, COMMD8, MYO1G, SP140, PHPT1, CD2BP2, PLEKHO1,
 STAM, MRPL16, IL2RB, ID2, TESPA1, GOLGA8B, MIS18BP1, VAMP5, DAPK2, HLA-

DPA1, TSG101, IL4R, CCND2, CTSC, TRAF3IP3, NLRC3, ORAI3, GNLY, MIR155HG, CARD16, CD82, ECH1, JAML, EEF1G, ETFB, DAXX, RBM4, HCST, RAB27A, YPEL2, CHST12, ARPC1B, PDIA4, PDIA6, AC243960.1, TBC1D10C, PTPN6, PYCARD, BST2, BTN3A2, MTG1, MLEC, DUSP4, GSDMD, SLAMF1, IFI6, PCID2, GIMAP1, ITGA1, CSNK2B, CDK2AP2, MYO1F, AC004687.1, PTTG1, APOBEC3C, TSPAN14, MOB3A, STXBP2, LCP2, PLA2G16, LINC00649, CST7, TADA3, SIT1, APOBEC3G, SUSP3, CD3G, CCL5, CDC25B, TNFRSF1B, HMGN3, THEMIS, ASF1A, CTNNA1, FIBP, CCDC85B, POLR3GL, GIMAP6, ARL6IP1, CALCOCO2, CCPG1, KLRB1, ACAA2, ISG15, EIF4A1, CAT, MANF, XAB2, GRINA, GLO1, LSM2, SLFN5, FKBP1A, AKNA, TAP1, LMO4, APEH, C12orf75, TMEM14A, DNPH1, C17orf49, NUDT5, MGAT1, CCDC69, EIF4EBP1, PDHB, ARL3, UCP2, IFI35, HSBP1, LYST, MRFAP1L1, ITGAL, AIP, RASAL3, CAPN1, ITGB1, RBPJ, LBH, DYNLL1, NME2, MT1F, SYNGR2, ABTB1, ZGPAT, CD63, ILK, SKA2, TMEM204, ACO2, HOPX, CRIP1, OXNAD1, CCS, GRAP2, GSTO1, HADHB, IL16, PIN4, CUEDC2, CALM3, SAMSNA1, HM13, SNAP23, LPCAT4, FAAP20, EFHD2, PRDX3, CCM2, C22orf39, SDHA, ARDC1, MAP4K1, NDUFA13, IL27RA, and C14orf119. These genes were used for Examples 8.1 and 8.3.

Genes used in Lowery et. al. (2022) (CD4):

CXCL13, HMOX1, ETV7, ADGRG1, PDCD1, ENTPD1, CCDC50, TOX, CD4, TIGIT, TNFRSF18, NMB, MYL6B, AHI1, MAF, IFNG, LAG3, CXCR6, IGFLR1, DUSP4, ACP5, LINC01943, LIMS1, BATF, PCED1B, ITGAL, YPEL2, MAL, PPT1, ELMO1, MIS18BP1, TMEM173, ADI1, SLA, GALM, LBH, SECISBP2L, CTSB, C17orf49, CORO1B, CARHSP1, SRPK2, ARL3, PTMS, CD82, HNRNP35, CTSC, LINC01871, CCDC167, SMC3, PPM1G, ORMDL3, VPS25, BST2, TRAF3IP3, NAP1L4, HLA-DPA1, PIM2, SH2D1A, RILPL2, and CCNDBP1. These genes were used for Example 8.2.

Genes used in WO 2021/188954 A1 (CD8)

AFAP11L2, ASB2, CXCL13, HMOX1, ITM2A, KLRB1, PDLIM4, TIGIT, AFAP11L2, ALOX5AP, ARHGAP9, ASB2, CARD16, CD3G, CD8A, CD8B, CLIC3, CTSW, CXCL13, CXCR6, GALNT2, GZMB, HLA-PDA1, HLA-DPB1, HLA-DRB1, HLA-DRB5, HMGN3, HMOX1, ITGAE, ITM2A, KLRB1, MPST, NAP1L4, NELL2, NSMCE1, PDLIM4, PTMS, RAB27A, RARRES3, RBPJ, TIGIT, CD39, CD74, CD103, CD106, CD137, HLA-DR, TIGIT, ABI3, AC243960.1, ACP5, ADGRG1, AHI1, ASB2, BST2, CARS, CCL4, CD27, CD2BP2,

CD82, CTSW, CXCL13, CXCR6, DUSP4, ENTPD1, GALNT2, GATA3, GPR25, GZMB, HDLBP, HLA-DPA1, HLA-DRB1, HMOX1, ID2, IGFLR1, LINC01871, LINC01943, MIS18BP1, MPST, NCF4, NSMCE1, PCED1B, PDCD1, PHPT1, PLEKHF1, PRF1, PTMS, SLC1A4, SLF1, SMC4, SUPT3H, TIGIT, TNFRSF18, TOX, TRAF3IP3, YPEL2, 5 AC243829.4, ACP5, APOBEC3C, APOBEC3G, CCL3, CCL4, CCL4L2, CCL5, CD27, CD8A, CD8B, CST7, CTSW, CXCL13, DUSP4, ENTPD1, FABP5, GALNT2, GNLY, GZMA, GZMB, GZMH, GZMK, HAVCR2, HCST, HLA-DMA, HLA-DPA1, HLA-DPB1, HLA-DRA, HLA-DRB1, HLA-DRB5, HMOX1, IFNG, IGFLR1, ITGAL, JAML, LINC01871, LYST, MIR155HG, NKG7, PLEKHF1, PRF1, PTMS, RGS1, SLF1, SMC4, 10 SUPT3H, TIGIT, TOX, AHI1, CXCL13, FABP5, NAP1L4, ORMDL3, PPP1R116B, SH2D1A, TIGIT, TOX, TIGIT, CD39, PD-1, LTB, LYAR, RGCC, S100A10, CD39, CD74, CD103, CD106, CD137, HLA-DR, TIGIT, CCR7, CD8A, CD16, CD45RA, CD62L, and IL7R. These genes were used for Example 8.1 and 8.3.

15 Genes used in WO 2021/188954 A1 (CD4)
AFAP11L2, ASB2, CXCL13, HMOX1, ITM2A, KLRB1, PDLIM4, TIGIT, BATF, CD247, DNPH1, DUSP4, GYPC, IFITM1, IGFLR1, LIMS1, NMB, NR3C1, SH2D1A, SPOCK2, SUPT3H, TNFRSF18, ADI1, AHI1, AR1D5B, CMTM7, CPM, CYTH1, ELMO1, ETV7, FABP5, FBLN7, FKBP5, GRAMD1A, HIF1A, IL6ST, ITGA4, ITK, JAK3, LEF1, MAF, 20 MAL, MIR4435-2HG, MYL6B, NAP1L4, PASK, PGM2L1, PIM2, PPP1CC, SESN3, SOCS1, STAT1, SYNE2, TBC1D4, TLK1, TMEM123, TMEM70, TNK1, TOX, TSHZ2, UCP2, VOPP1, YPEL2, ABI3, AC243960.1, ACP5, ADGRG1, BST2, CARS, CCL4, CD27, CD2BP2, CD82, CTSW, CXCR6, ENTPD1, GALNT2, GATA3, GPR25, GZMB, HDLBP, HLA-DPA1, HLA-DRB1, ID2, LINC01871, LINC01943, MIS18BP1, MPST, NCF4, 25 NSMCE1, PCED1B, PDCD1, PHPT1, PLEKHF1, PRF1, PTMS, SLC1A4, SLF1, SMC4, TRAF3IP3, ORMDL3, PPP1R116B, CD39, PD-1, LTB, LYAR, RGCC, S100A10, CCL5, CD52, GSTSP1, JUN, LGALS1, PLP2, VIM, and ZFP36. These genes were used for Example 8.2.

References cited:

- Caushi *et al.* (2021), Nature 596(7870):126
- Cano-Gamez et al. (2020), Nat Comm 11:, art. 1801 (doi.org/10.1038/s41467-020-15543-y)
- 5 Iwabuchi & van Kaer (2019), Front Immunol 10:1837 (doi: 10.3389/fimmu.2019.01837)
- Lowery et al. (2022), Science 10.1126/science.abl5447
- Magen et al. (2019), Cell Rep 29(10):3019 (doi.org/10.1016/j.celrep.2019.10.131)
- MacQueen (1967), "Some methods for classification and analysis of multivariate observations", 5th Berkeley Symposium on Mathematical Statistics and Probability
- 10 McInnes et al. (2020), arXiv:1802.03426v3
- Oh et al. (2020), Cell 181(7):1612 (doi.org/10.1016/j.cell.2020.05.017)
- van der Maaten and Hinton (2008), J Machine Learning Res 9:2579
- WO2018/209324
- WO2019/070755
- 15 WO 2021/188954 A1

Table 1: Biomarkers of the core signature

No.	Gene.Name	Signature	Full. Gene.Name	ENSEMBL ID	ENTREZ ID	Refseq.version
1	CCL4	Core	C-C motif chemokine ligand 4	ENSG00000275302	6351	NM_002984.4
2	CCL4L2	Core	C-C motif chemokine ligand 4 like 2	ENSG00000276070	9560	NM_001001435.2
3	CCL3	Core	C-C motif chemokine ligand 3	ENSG00000277632	6348	NM_002983.3
4	CCL3L1	Core	C-C motif chemokine ligand 3 like 1	ENSG00000277796	6349	NM_021006.5
5	CXCL13	Core	C-X-C motif chemokine ligand 13	ENSG00000156234	10563	NM_001371558.1

Table 2: Biomarkers of the accessory 1 signature

No.	Gene.Name	Signature	Full. Gene.Name	ENSEMBL ID	ENTREZ ID	Refseq.version
6	IFNG	Accessory 1	interferon gamma(IFNG)	ENSG00000111537	3458	NM_000619.3
7	HAVCR2	Accessory 1	hepatitis A virus cellular receptor 2(HAVCR2)	ENSG00000135077	84868	NM_032782.5
8	FBNP1	Accessory 1	formin binding protein 1(FBNP1)	ENSG00000187239	23048	NM_001363755.1
9	CSRNP1	Accessory 1	cysteine and serine rich nuclear protein 1(CSRNP1)	ENSG00000144655	64651	NM_001320559.2
10	SPRY1	Accessory 1	sprouty RTK signaling antagonist 1(SPRY1)	ENSG00000164056	10252	NM_001258038.2
11	RHOH	Accessory 1	ras homolog family member H(RHOH)	ENSG00000168421	399	NM_001278359.2
12	FOXN2	Accessory 1	forkhead box N2(FOXN2)	ENSG00000170802	3344	NM_002158.4
13	HIF1A	Accessory 1	hypoxia inducible factor 1 alpha subunit(HIF1A)	ENSG00000100644	3091	NM_001243084.2

14	TOB1	Accessory 1	transducer of ERBB2 1 (TOB1)	ENSG000000141232	10140	NM_001243877.2
15	RILPL2	Accessory 1	Rab interacting lysosomal protein like 2(RILPL2)	ENSG000000150977	196383	NM_145058.3
16	CD8B	Accessory 1	CD8b molecule(CD8B)	ENSG000000172116	926	NM_001178100.2
17	GABARAPL1	Accessory 1	GABA type A receptor associated protein like 1(GABARAPL1)	ENSG000000139112	23710	NM_001363598.2
18	TNFSF14	Accessory 1	tumor necrosis factor superfamily member 14(TNFSF14)	ENSG000000125735	8740	NM_003807.5
19	EGR1	Accessory 1	early growth response 1(EGR1)	ENSG000000120738	1958	NM_001964.3
20	EGR2	Accessory 1	early growth response 2(EGR2)	ENSG000000122877	1959	NM_000399.5
21	TAGAP	Accessory 1	T-cell activation RhoGTPase activating protein(TAGAP)	ENSG000000164691	117289	NM_001278733.2
22	TNFSF9	Accessory 1	tumor necrosis factor superfamily member 9(TNFSF9)	ENSG000000125657	8744	NM_003811.4
23	ANXA1	Accessory 1	annexin A1(ANXA1)	ENSG000000135046	301	NM_000700.3
24	MAP3K8	Accessory 1	mitogen-activated protein kinase kinase kinase 8(MAP3K8)	ENSG000000107968	1326	NM_001244134.1
25	PIK3R1	Accessory 1	phosphoinositide-3-kinase regulatory subunit 1(PIK3R1)	ENSG000000145675	5295	NM_001242466.2
26	DUSP2	Accessory 1	dual specificity phosphatase 2(DUSP2)	ENSG000000158050	1844	NM_004418.4
27	DUSP4	Accessory 1	dual specificity phosphatase 4(DUSP4)	ENSG000000120875	1846	NM_001394.7
28	DUSP6	Accessory 1	dual specificity phosphatase 6(DUSP6)	ENSG000000139318	1848	NM_001946.4
29	CLIC3	Accessory 1	chloride intracellular channel 3(CLIC3)	ENSG000000169583	9022	NM_004669.3

30	RASGEF1B	Accessory 1	RasGEF domain family member 1B(RASGEF1B)	ENSG000000138670	153020	NM_001300735.2
31	LAG3	Accessory 1	lymphocyte activating 3(LAG3)	ENSG000000089692	3902	NM_002286.6
32	XCL2	Accessory 1	X-C motif chemokine ligand 2(XCL2)	ENSG000000143185	6846	NM_003175.4
33	NR4A2	Accessory 1	nuclear receptor subfamily 4 group A member 2(NR4A2)	ENSG000000153234	4929	NM_006186.4
34	DNAJB6	Accessory 1	DnaJ heat shock protein family (Hsp40) member B6(DNAJB6)	ENSG000000105993	10049	NM_001363676.1
35	NFKBID	Accessory 1	NFKB inhibitor delta(NFKBID)	ENSG000000167604	84807	NM_001321831.2
36	MCL1	Accessory 1	BCL2 family apoptosis regulator(MCL1)	ENSG000000143384	4170	NM_001197320.2
37	EVI2A	Accessory 1	ecotropic viral integration site 2A(EVI2A)	ENSG000000126860	2123	NM_001003927.3
38	SLC7A5	Accessory 1	solute carrier family 7 member 5(SLC7A5)	ENSG000000103257	8140	NM_003486.7
39	H3F3B	Accessory 1	H3 histone family member 3B(H3F3B)	ENSG000000132475	3021	NM_005324.5
40	NR4A3	Accessory 1	nuclear receptor subfamily 4 group A member 3(NR4A3)	ENSG000000119508	8013	NM_006981.4
41	REL	Accessory 1	REL proto-oncogene NF-kB subunit(REL)	ENSG000000162924	5966	NM_001291746.2
42	IRF4	Accessory 1	interferon regulatory factor 4(IRF4)	ENSG000000137265	3662	NM_001195286.2
43	CST7	Accessory 1	cystatin F(CST7)	ENSG000000077984	8530	NM_003650.4
44	ATF3	Accessory 1	activating transcription factor 3(ATF3)	ENSG000000162772	467	NM_001030287.4
45	TNF	Accessory 1	tumor necrosis factor(TNF)	ENSG000000232810	7124	NM_000594.4
46	GPR171	Accessory 1	G protein-coupled receptor 171(GPR171)	ENSG000000174946	29909	NM_013308.4
47	BCL2A1	Accessory 1	BCL2 related protein A1(BCL2A1)	ENSG000000140379	597	NM_001114735.2
48	ITGA1	Accessory 1	integrin subunit alpha 1(ITGA1)	ENSG000000213949	3672	NM_181501.2

49	TNFAIP3	Accessory 1	TNF alpha induced protein 3(TNFAIP3)	ENSG000000118503	7128	NM_001270507.2
50	NR4A1	Accessory 1	nuclear receptor subfamily 4 group A member 1(NR4A1)	ENSG000000123358	3164	NM_001202233.2
51	RUNX3	Accessory 1	runx related transcription factor 3(RUNX3)	ENSG000000020633	864	NM_001031680.2
52	HERPUD2	Accessory 1	HERPUD family member 2(HERPUD2)	ENSG000000122557	64224	NM_022373.5
53	FASLG	Accessory 1	Fas ligand(FASLG)	ENSG000000117560	356	NM_000639.3
54	CBLB	Accessory 1	Cbl proto-oncogene B(CBLB)	ENSG000000114423	868	NM_001321786.1
55	PTGER4	Accessory 1	prostaglandin E receptor 4(PTGER4)	ENSG000000171522	5734	NM_000958.3
56	SLA	Accessory 1	Src-like-adaptor(SLA)	ENSG000000155926	6503	NM_001045556.3
57	XCL1	Accessory 1	X-C motif chemokine ligand 1(XCL1)	ENSG000000143184	6375	NM_002995.3
58	BHLHE40	Accessory 1	basic helix-loop-helix family member e40(BHLHE40)	ENSG000000134107	8553	NM_003670.3
59	LYST	Accessory 1	lysosomal trafficking regulator(LYST)	ENSG000000143669	1130	NM_000081.4
60	KLRD1	Accessory 1	killer cell lectin like receptor D1(KLRD1)	ENSG000000134539	3824	NM_001114396.3
61	ZNF682	Accessory 1	zinc finger protein 682(ZNF682)	ENSG000000197124	91120	NM_001077349.1
62	CTSW	Accessory 1	cathepsin W(CTSW)	ENSG000000172543	1521	NM_001335.4
63	SLC2A3	Accessory 1	solute carrier family 2 member 3(SLC2A3)	ENSG000000059804	6515	NM_006931.3
64	NLRP3	Accessory 1	NLR family pyrin domain containing 3(NLRP3)	ENSG000000162711	114548	NM_001079821.3
65	SCML4	Accessory 1	sex comb on midleg-like 4 (Drosophila)(SCML4)	ENSG000000146285	256380	NM_001286408.2
66	VSIR	Accessory 1	V-Set Immunoregulator Receptor (VSIR)	ENSG000000107738	64115	NM_022153.2

67	LINC01871	Accessory 1	Long Intergenic Non-Protein Coding RNA 1871 (LINC01871)	ENSG00000235576	101929531	XR_001739273.1
68	ZFP36L1	Accessory 1	ZFP36 Ring Finger Protein Like 1	ENSG00000185650	677	NM_001244698.2

Table 3: Biomarkers of the accessory 2 signature

No.	Gene.Name	Signature	Full.Gene.Name	ENSEMBL ID	ENTREZ ID	Refseq.version
69	CCL5	Accessory 2	C-C motif chemokine ligand 5(CCL5)	ENSG00000271503	6352	NM_001278736.2
70	GZMH	Accessory 2	granzyme H(GZMH)	ENSG00000100450	2999	NM_001270780.2
71	CLEC2B	Accessory 2	C-type lectin domain family 2 member B(CLEC2B)	ENSG00000110852	9976	NM_005127.3
72	GZMA	Accessory 2	granzyme A(GZMA)	ENSG00000145649	3001	NM_006144.4
73	CD69	Accessory 2	CD69 molecule(CD69)	ENSG00000110848	969	NM_001781.2
74	GZMK	Accessory 2	granzyme K(GZMK)	ENSG00000113088	3003	NM_002104.3
75	CRTAM	Accessory 2	cytotoxic and regulatory T-cell molecule(CRTAM)	ENSG00000109943	56253	NM_001304782.2

5 Table 4: Biomarkers of the exclusion signature

No.	Gene.Name	Signature	Full.Gene.Name	ENSEMBL ID	ENTREZ ID	Refseq.version
76	GNLY	Exclusion	granulysin	ENSG00000115523	10578	NM_001302758.2
77	FGFBP2	Exclusion	fibroblast growth factor binding protein 2	ENSG00000137441	83888	NM_031950.4

Table 5: Biomarkers of the core-2 signature in addition to CCL3L1

No.	Gene.Name	Signature	Full.Gene.Name	ENSEMBL ID	ENTREZ ID	Refseq.version
31	LAG3	core-2	lymphocyte activating 3(LAG3)	ENSG000000089692	3902	NM_002286.6
17	GA-BARAPL1	core-2	GABA type A receptor associated protein like 1(GABARAPL1)	ENSG000000139112	23710	NM_001363598.2
54	CBLB	core-2	Cbl proto-oncogene B(CBLB)	ENSG000000114423	868	NM_001321786.1
56	SLA	core-2	Src-like-adaptor(SLA)	ENSG000000155926	6503	NM_001045556.3
60	KLRD1	core-2	killer cell lectin like receptor D1(KLRD1)	ENSG000000134539	3824	NM_001114396.3
71	CLEC2B	core-2	C-type lectin domain family 2 member B(CLEC2B)	ENSG000000110852	9976	NM_005127.3

Table 6: Biomarkers of signature 2

No.	Gene.Name	Signature	Full.Gene.Name	ENSEMBL ID	ENTREZ ID	Refseq.version
78	CTSD	Signature 2	cathepsin D	ENSG000000117984	1509	NM_001909.5
79	CD7	Signature 2	CD7 molecule	ENSG000000173762	924	NM_006137.7
80	CD3D	Signature 2	CD3d molecule	ENSG000000167286	915	NM_000732.6

81	LSP1	Signature 2	lymphocyte specific protein 1	ENSG00000130592	4046	NM_001013253.2
82	SNAP47	Signature 2	synaptosome associated protein 47	ENSG00000143740	116841	NM_001323930.2
83	GAPDH	Signature 2	glyceraldehyde-3-phosphate dehydrogenase	ENSG00000111640	2597	NM_001256799.3
84	KLRK1	Signature 2	killer cell lectin like receptor K1	ENSG00000213809	22914	NM_007360.4
85	TNS3	Signature 2	tensin 3	ENSG00000136205	64759	NM_022748.12
86	VCAM1	Signature 2	vascular cell adhesion molecule 1	ENSG00000162692	7412	NM_001078.4
87	KLRC2	Signature 2	killer cell lectin like receptor C2	ENSG00000205809	3822	NM_002260.4
88	PMAIP1	Signature 2	phorbol-12-myristate-13-acetate-induced protein 1	ENSG00000141682	5366	NM_001382615.1
89	FYN	Signature 2	FYN proto-oncogene, Src family tyrosine kinase	ENSG00000010810	2534	NM_001370529.1
90	CTLA4	Signature 2	cytotoxic T-lymphocyte associated protein 4	ENSG00000163599	1493	NM_001037631.3
91	GSTP1	Signature 2	glutathione S-transferase pi 1	ENSG00000084207	2950	NM_000852.4
92	AREG	Signature 2	amphiregulin	ENSG00000109321	374	NM_001657.4
93	FAM3C	Signature 2	FAM3 metabolism regulating signaling molecule C	ENSG00000196937	10447	NM_001040020.2
94	SH3BGRL3	Signature 2	SH3 domain binding glutamate rich protein like 3	ENSG00000142669	83442	NM_031286.4
95	CD3E	Signature 2	CD3e molecule	ENSG00000198851	916	NM_000733.4
96	SRGAP3	Signature 2	SLIT-ROBO Rho GTPase activating protein 3	ENSG00000196220	9901	NM_001033117.3
97	SRGN	Signature 2	serglycin	ENSG00000122862	5552	NM_001321053.2
98	SIRPG	Signature 2	signal regulatory protein gamma	ENSG00000089012	55423	NM_001039508.2
99	SCPEP1	Signature 2	serine carboxypeptidase 1	ENSG00000121064	59342	NM_021626.3

100	RHOB	Signature 2	ras homolog family member B	ENSG00000143878	388	NM_004040.4
101	ANKRD28	Signature 2	ankyrin repeat domain 28	ENSG00000206560	23243	NM_001195098.1
102	LINC02446	Signature 2	long intergenic non-protein coding RNA 2446	NA	101060038	NR_146455.1
103	RABAC1	Signature 2	Rab acceptor 1	ENSG00000105404	10567	NM_006423.3
104	IKZF3	Signature 2	IKAROS family zinc finger 3	ENSG00000161405	22806	NM_001257408.2
105	BCAS4	Signature 2	breast carcinoma amplified sequence 4	ENSG00000124243	55653	NM_001010974.2
106	CD2	Signature 2	CD2 molecule	ENSG00000116824	914	NM_001328609.2
107	BLOC1S1	Signature 2	biogenesis of lysosomal organelles complex 1 subunit 1	ENSG00000135441	2647	NM_001487.4
108	RHOA	Signature 2	ras homolog family member A	ENSG000000067560	387	NM_001313941.2
109	EID1	Signature 2	EP300 interacting inhibitor of differentiation 1	ENSG00000255302	23741	NM_014335.3
110	MYL6	Signature 2	myosin light chain 6	ENSG000000092841	4637	NM_021019.5
111	CLIC1	Signature 2	chloride intracellular channel 1	ENSG00000213719	1192	NM_001287593.1
112	IQGAP1	Signature 2	IQ motif containing GTPase activating protein 1	ENSG00000140575	8826	NM_003870.4
113	ARPC2	Signature 2	actin related protein 2/3 complex subunit 2	ENSG00000163466	10109	NM_005731.3
114	PHYKPL	Signature 2	5-phosphohydroxy-L-lysine phospho-lyase	ENSG00000175309	85007	NM_001278346.1
115	PRDM1	Signature 2	PR/SET domain 1	ENSG00000057657	639	NM_001198.4
116	EVL	Signature 2	Enah/Vasp-like	ENSG00000196405	51466	NM_001330221.2
117	TPI1	Signature 2	triosephosphate isomerase 1	ENSG00000111669	7167	NM_000365.6
118	ADGRE5	Signature 2	adhesion G protein-coupled receptor E5	ENSG00000123146	976	NM_001025160.3

119	PAXX	Signature 2	PAXX non-homologous end joining factor	ENSG000000148362	286257	NM_001329678.2
120	RGS2	Signature 2	regulator of G protein signaling 2	ENSG000000116741	5997	NM_002923.4
121	HERPUD1	Signature 2	homocysteine inducible ER protein with ubiquitin like domain 1	ENSG000000051108	9709	NM_001010989.3
122	IFI27L2	Signature 2	interferon alpha inducible protein 27 like 2	ENSG000000119632	83982	NM_032036.3
123	SEPTIN7	Signature 2	septin 7	ENSG000000122545	989	NM_001011553.4
124	UBB	Signature 2	ubiquitin B	ENSG000000170315	7314	NM_001281716.2
125	JUN	Signature 2	Jun proto-oncogene, AP-1 transcription factor subunit	ENSG000000177606	3725	NM_002228.4
126	CFLAR	Signature 2	CASP8 and FADD like apoptosis regulator	ENSG000000003402	8837	NM_001127183.4
127	LITAF	Signature 2	lipopolysaccharide induced TNF factor	ENSG000000189067	9516	NM_001136472.2
128	ANXA5	Signature 2	annexin A5	ENSG000000164111	308	NM_001154.4
129	STAT3	Signature 2	signal transducer and activator of transcription 3	ENSG000000168610	6774	NM_001369512.1
130	RSRP1	Signature 2	arginine and serine rich protein 1	ENSG000000117616	57035	NM_001321772.2
131	PRDX5	Signature 2	peroxiredoxin 5	ENSG000000126432	25824	NM_001358511.2
132	SEM1	Signature 2	SEM1 26S proteasome subunit	ENSG000000127922	7979	NM_001393898.1
133	SERPINB1	Signature 2	serpin family B member 1	ENSG000000021355	1992	NM_030666.4
134	RNF19A	Signature 2	ring finger protein 19A RBR E3 ubiquitin protein ligase	ENSG0000000034677	25897	NM_001280539.2
135	IL2RG	Signature 2	interleukin 2 receptor subunit gamma	ENSG000000147168	3561	NM_000206.3
136	ENSA	Signature 2	endosulfine alpha	ENSG000000143420	2029	NM_004436.4
137	SRP14	Signature 2	signal recognition particle 14	ENSG000000140319	6727	NM_001309434.1

138	ATP6V0C	Signature 2	ATPase H ⁺ transporting V0 subunit c	ENSG000000185883	527	NM_001198569.2
139	LY6E	Signature 2	lymphocyte antigen 6 family member E	ENSG000000160932	4061	NM_001127213.2
140	BIN1	Signature 2	bridging integrator 1	ENSG000000136717	274	NM_001320632.2
141	AKAP13	Signature 2	A-kinase anchoring protein 13	ENSG000000170776	11214	NM_001270546.1
142	PDE4D	Signature 2	phosphodiesterase 4D	ENSG000000113448	5144	NM_001104631.2
143	PELI1	Signature 2	pellino E3 ubiquitin protein ligase 1	ENSG000000197329	57162	NM_020651.4
144	PARK7	Signature 2	Parkinsonism associated deglycase	ENSG000000116288	11315	NM_001123377.2
145	MSN	Signature 2	moesin	ENSG000000147065	4478	NM_002444.3
146	SERTAD1	Signature 2	SERTA domain containing 1	ENSG000000197019	29950	NM_013376.4
147	RAC2	Signature 2	Rac family small GTPase 2	ENSG000000128340	5880	NM_002872.5
148	SELENOH	Signature 2	selenoprotein H	ENSG000000211450	280636	NM_001321335.2
149	PSMB8	Signature 2	proteasome 20S subunit beta 8	ENSG000000204264	5696	NM_004159.5
150	CKLF	Signature 2	chemokine like factor	ENSG000000217555	51192	NM_001040138.3
151	KLRC1	Signature 2	killer cell lectin like receptor C1	ENSG000000134545	3821	NM_001304448.1
152	RNASEK	Signature 2	ribonuclease K	ENSG000000219200	440400	NM_001004333.5
153	MT2A	Signature 2	metallothionein 2A	ENSG000000125148	4502	NM_005953.5
154	TXNIP	Signature 2	thioredoxin interacting protein	ENSG000000265972	10628	NM_001313972.2
155	CD40LG	Signature 2	CD40 ligand	ENSG000000102245	959	NM_000074.3
156	DRAIC	Signature 2	downregulated RNA in cancer; inhibitor of cell invasion and migration	ENSG000000245750	145837	NR_026979.1
157	FOXP3	Signature 2	forkhead box P3	ENSG000000049768	50943	NM_001114377.2

Table 7: Exemplary F-score values obtainable with markers and their combinations for non-primary brain metastasis and lung cancer

Signature	F-Score brain metastasis	F-score lung cancer
CCL4	0,4798	0,9481
Comparative Example B	0,5296	0,10667
Comparative Example A	0,7031	0,9481
CCL3	0,7064	0,9559
CCL3L1 + CCL4 + CCL4L2 + LAG3 + KLRD1	0,7081	0,9559
CCL4L2 + CCL3	0,7095	0,971
CCL4 + CCL4L2 + LAG3 + KLRD1	0,7115	0,9929
CCL4 + CCL3	0,7116	0,9929
CCL4 + CCL3 + LAG3 + KLRD1	0,7117	0,9857
CCL3L1 + CCL4 + CCL3 + LAG3 + KLRD1	0,7117	0,9857
CCL3L1 + CCL3	0,7151	0,9784
CCL3L1 + CCL4L2 + CCL3 + LAG3 + KLRD1	0,7156	0,9857
CCL3L1 + CCL4 + KLRD1	0,7167	0,9929
CCL3L1 + CCL4 + CCL4L2 + CCL3 + KLRD1	0,7178	0,9857
CCL4 + CCL4L2 + CCL3 + LAG3	0,7179	0,9784
CCL3L1 + CCL4 + LAG3 + KLRD1	0,719	0,9559
CCL4 + CCL4L2 + KLRD1	0,7197	0,971
CCL3L1 + CCL4L2 + LAG3	0,7198	0,971
KLRD1 + LAG3	0,7227	1
CCL3L1 + CCL4 + CCL3	0,7258	0,9635
CCL3L1 + CCL3 + KLRD1	0,7273	0,9929
CCL3L1 + CCL4L2 + LAG3 + KLRD1	0,7277	0,971
CCL3L1 + CCL4L2 + CCL3	0,7283	0,9559
CCL4L2 + LAG3 + KLRD1	0,7306	0,971
CCL4L2 + CCL3 + KLRD1	0,7326	1
KLRD1 + CCL3	0,7368	0,9929
CCL3L1 + CCL4 + CCL3 + LAG3	0,7368	0,9929
CCL3L1 + CCL3 + LAG3	0,7379	0,9857

CCL3L1 + LAG3 + KLRD1	0,7385	0,9784
CCL4 + CCL3 + KLRD1	0,7442	0,9784
CCL3L1 + CCL4L2 + KLRD1	0,7445	0,971
CCL4 + LAG3 + KLRD1	0,7461	0,9784
CCL4 + CCL4L2 + CCL3 + KLRD1	0,7549	0,971
CCL3L1 + CCL4L2 + CCL3 + KLRD1	0,7566	0,9857
KLRD1	0,757	0,9559
KLRD1 + CCL4L2	0,7643	0,9635
CCL3L1 + CCL3 + LAG3 + KLRD1	0,7709	0,9784
CCL3L1 + CCL4 + CXCL13 + LAG3	0,7723	0,9857
CCL3L1 + CCL4 + CCL3 + CXCL13 + LAG3	0,7772	0,9784
CCL3 + LAG3 + KLRD1	0,7773	0,9929
CCL3L1 + CCL4 + CCL3 + CXCL13	0,7805	0,9929
KLRD1 + CCL3L1	0,7828	0,9635
CCL4L2 + CCL3 + LAG3 + KLRD1	0,7852	0,9857
CCL4 + CCL4L2 + CCL3 + CXCL13 + KLRD1	0,7894	0,9929
CCL4 + CCL3 + CXCL13 + LAG3	0,7916	0,9929
7 Alternative Core Genes	0,7917	0,9929
5 Core Genes	0,7921	0,9929
CCL3L1 + CCL4 + CCL4L2 + CXCL13 + LAG3	0,7921	1
CCL3L1 + CCL4 + CCL4L2 + CXCL13	0,793	0,9857
CCL3L1 + CCL4 + CCL4L2 + CXCL13 + KLRD1	0,7952	1
CCL3L1 + CCL4 + CCL3 + CXCL13 + LAG3 + KLRD1	0,7956	1
CCL3L1 + CCL4 + CCL4L2 + CCL3 + CXCL13 +	0,796	1
CCL4 + CCL3 + CXCL13 + KLRD1	0,7969	0,9929
CCL3L1 + CCL4 + CXCL13 + LAG3 + KLRD1	0,7983	1
CCL4 + CXCL13 + LAG3	0,8009	0,9857
Signature 2	0,8022	0,9857
CCL3L1 + CCL4 + CCL4L2 + CCL3 + CXCL13 +	0,803	0,9929
CCL3L1 + CCL4 + CCL4L2 + CCL3 + CXCL13 +	0,8039	0,9929
CCL3L1 + CCL4L2 + CCL3 + CXCL13 + LAG3 +	0,8052	0,9929

CCL4 + CXCL13	0,8053	1
CCL4 + CCL4L2 + CCL3 + CXCL13	0,8065	0,9929
CCL4 + CCL4L2 + CXCL13	0,8081	0,9929
CCL3L1 + CCL4 + CXCL13 + KLRD1	0,8092	0,9929
CCL4 + CCL4L2 + CXCL13 + LAG3	0,8092	0,9929
CCL4 + CCL4L2 + CCL3 + CXCL13 + LAG3	0,8093	0,9857
CCL3L1 + CXCL13 + LAG3 + KLRD1	0,81	0,9857
CCL3L1 + CCL4 + CCL4L2 + CXCL13 + LAG3 +	0,8112	0,9929
CCL4 + CCL4L2 + CCL3 + CXCL13 + LAG3 + KLRD1	0,8129	1
CCL4 + CCL4L2 + CXCL13 + LAG3 + KLRD1	0,8134	0,9929
CCL4L2 + CCL3 + CXCL13 + LAG3 + KLRD1	0,8159	1
CCL3L1 + CCL4 + CXCL13	0,8165	0,9857
CCL4 + CCL4L2 + CXCL13 + KLRD1	0,8186	1
CCL3L1 + CCL3 + CXCL13 + KLRD1	0,8206	0,9784
CCL4 + CXCL13 + KLRD1	0,8222	1
CXCL13 + LAG3 + KLRD1	0,8228	1
CCL4 + CCL3 + CXCL13 + LAG3 + KLRD1	0,8234	0,9784
CCL3L1 + CCL4L2 + CXCL13	0,8239	0,9857
CCL3L1 + CCL4L2 + CXCL13 + KLRD1	0,8242	0,9929
CCL4L2 + CXCL13 + LAG3 + KLRD1	0,8261	0,9857
CCL4 + CCL3 + CXCL13	0,8272	0,9929
CCL3L1 + CCL4L2 + CXCL13 + LAG3	0,8274	0,9784
CCL3L1 + CXCL13 + KLRD1	0,8278	0,9857
CCL3L1 + CCL4L2 + CCL3 + CXCL13 + LAG3	0,8295	0,9929
CCL3 + CXCL13 + KLRD1	0,8295	1
CCL4 + CXCL13 + LAG3 + KLRD1	0,8298	1
KLRD1 + CXCL13	0,8326	0,9857
CCL3L1 + CCL3 + CXCL13 + LAG3	0,8326	1
CCL4L2 + CCL3 + CXCL13	0,8328	0,9857
CCL3L1 + CCL4L2 + CCL3 + CXCL13 + KLRD1	0,834	0,9857
CCL3L1 + CCL4L2 + CCL3 + CXCL13	0,8349	0,9784

CCL3L1 + CXCL13 + LAG3	0,8374	0,971
CCL3 + CXCL13	0,8387	0,9784
CCL3 + CXCL13 + LAG3	0,84	1
CCL4L2 + CXCL13 + LAG3	0,8408	0,9784
CCL4L2 + CCL3 + CXCL13 + KLRD1	0,841	0,9929
CCL4L2 + CXCL13 + KLRD1	0,8418	0,9784
CCL3L1 + CCL3 + CXCL13	0,842	0,9635
CCL3L1 + CXCL13	0,8421	0,9784
CCL3 + CXCL13 + LAG3 + KLRD1	0,8433	0,9929
CCL3L1 + CCL3 + CXCL13 + LAG3 + KLRD1	0,8465	0,9929
CCL4L2 + CXCL13	0,848	0,9857
CCL3L1 + CCL4L2 + CXCL13 + LAG3 + KLRD1	0,8492	0,9857
CCL4L2 + CCL3 + CXCL13 + LAG3	0,8493	0,9857
CXCL13 + LAG3	0,8515	1
CCL3L1 + CCL4 + CCL3 + CXCL13 + KLRD1	0,8554	0,9929

Table 8: Exemplary additional F-score values obtainable with markers and their combinations for non-primary brain metastasis

Signature	F-Score
CCL3L1	0,7126
Core + Accessory Genes	0,7935
CXCL13	0,8463

Table 9: Exemplary additional F-score values obtainable with markers and their combinations for lung cancer

Signature	F-Score
CCL3 + LAG3	0,9857
CCL3L1 + CCL4 + CCL3 + KLRD1	0,9857
CCL3L1 + CCL4 + CCL4L2 + CCL3	0,971
CCL3L1 + CCL4 + CCL4L2 + CCL3 + CXCL13 +	1
CCL3L1 + CCL4 + CCL4L2 + CCL3 + LAG3	0,9857

CCL3L1 + CCL4 + CCL4L2 + LAG3	0,9784
CCL3L1 + CCL4 + CCL4L2 + LAG3 + KLRD1	0,9559
CCL3L1 + CCL4 + LAG3	0,9635
CCL3L1 + CCL4L2 + CCL3 + LAG3	0,9784
CCL4 + CCL3 + LAG3	0,9857
CCL4 + CCL4L2	0,971
CCL4 + CCL4L2 + CCL3	0,9857
CCL4 + CCL4L2 + CCL3 + LAG3 + KLRD1	0,9784
CCL4 + CCL4L2 + LAG3	0,9929
CCL4 + LAG3	1
KLRD1 + CCL4	0,971
LAG3	0,9559

Table 10: Exemplary F-score values obtainable with markers and their combinations for glioma

Signature	F-Score
Comparative Example B	0,12308
Comparative Example A	0,5227
Signature 2	0,5333
7 Alternative Core Genes + CD8B	0,8361
Core + Accessory Genes +CD8B	0,86179
5 Core Genes + CD8B	0,864
Signature 2 + CD8B	0,8644

Claims

1. A method of identifying a T-cell reactive to cells of a subject presenting a T-cell activating antigen (reactive T-cell), comprising
 - (a) determining expression of at least one of CCL3L1, CCL4, CCL4L2, CCL3, and CXCL13 in T-cells from a sample of said subject; and
 - (b) identifying a reactive T-cell based on the determination of step (a), preferably wherein said T-cell activating antigen is a cancer antigen or an autoimmune T-cell activating antigen, more preferably is a cancer antigen.
2. The method of claim 1, wherein step (a) comprises determining expression of at least two, preferably at least three, more preferably at least four, of CCL3L1, CCL4, CCL4L2, CCL3, and CXCL13.
3. The method of claim 1 or 2, wherein step (a) comprises further determining expression of at least one biomarker selected from the list consisting of IFNG, HAVCR2, FNBP1, CSRNPI, SPRY1, RHOH, FOXN2, HIF1A, TOB1, RILPL2, CD8B, GABARAPL1, TNFSF14, EGR1, EGR2, TAGAP, TNFSF9, ANXA1, MAP3K8, PIK3R1, DUSP2, DUSP4, DUSP6, CLIC3, RASGEF1B, LAG3, XCL2, NR4A2, DNAJB6, NFKBID, MCL1, EVI2A, SLC7A5, H3F3B, NR4A3, REL, IRF4, CST7, ATF3, TNF, GPR171, BCL2A1, ITGA1, TNFAIP3, NR4A1, RUNX3, HERPUD2, FASLG, CBLB, PTGER4, SLA, XCL1, BHLHE40, LYST, KLRD1, ZNF682, CTSW, SLC2A3, NLRP3, SCML4, VSIR, LINC01871, and ZFP36L1.
4. The method of any one of claims 1 to 3, wherein step (a) comprises further determining expression of at least one biomarker selected from the list consisting of LAG3, GABARAPL1, CBLB, SLA, KLRD1, and CLEC2B, preferably comprises determining all biomarkers of claim 1 and/or of claim 4.
5. The method of any one of claims 1 to 4, wherein step (a) comprises further determining expression of at least one biomarker selected from the list consisting of CTSD, CD7, CD3D, LSP1, SNAP47, GAPDH, KLRK1, TNS3, VCAM1, KLRC2, PMAIP1, FYN, CTLA4, GSTP1, AREG, FAM3C, SH3BGRL3, CD3E, SRGAP3, SRGN, SIRPG, SCPEP1, RHOB, ANKRD28, LINC02446, RABAC1, IKZF3, BCAS4, CD2,

BLOC1S1, RHOA, EID1, MYL6, CLIC1, IQGAP1, ARPC2, PHYKPL, PRDM1, EVL, TPI1, ADGRE5, PAXX, RGS2, HERPUD1, IFI27L2, SEPTIN7, UBB, JUN, CFLAR, LITAF, ANXA5, STAT3, RSRP1, PRDX5, SEM1, SERPINB1, RNF19A, IL2RG, ENSA, SRP14, ATP6V0C, LY6E, BIN1, AKAP13, PDE4D, PELI1, PARK7, MSN, SERTAD1, RAC2, SELENOH, PSMB8, CKLF, KLRC1, RNASEK, MT2A, TXNIP, and FOXP3.

6. The method of any one of claims 1 to 5, wherein step (a) comprises determining expression of at least CCL3L1 + CCL3; CCL3L1 + CCL3 + LAG3 + KLRD1; CCL3L1 + CCL3 + CXCL13; CCL3L1 + CCL3 + CXCL13 + KLRD1; CCL3L1 + CCL3 + CXCL13 + LAG3; CCL3L1 + CCL3 + CXCL13 + LAG3 + KLRD1; CCL3L1 + CCL3 + KLRD1; CCL3L1 + CCL3 + LAG3; CCL3L1 + CCL4 + CCL3; CCL3L1 + CCL4 + CCL3 + CXCL13; CCL3L1 + CCL4 + CCL3 + CXCL13 + KLRD1; CCL3L1 + CCL4 + CCL3 + CXCL13 + LAG3; CCL3L1 + CCL4 + CCL3 + CXCL13 + LAG3 + KLRD1; CCL3L1 + CCL4 + CCL3 + LAG3; CCL3L1 + CCL4 + CCL3 + LAG3 + KLRD1; CCL3L1 + CCL4 + CCL4L2 + CCL3 + CXCL13 + KLRD1; CCL3L1 + CCL4 + CCL4L2 + CCL3 + CXCL13 + LAG3 ; CCL3L1 + CCL4 + CCL4L2 + CCL3 + CXCL13 + LAG3 + KLRD1; CCL3L1 + CCL4 + CCL4L2 + CCL3 + KLRD1; CCL3L1 + CCL4 + CCL4L2 + CXCL13; CCL3L1 + CCL4 + CCL4L2 + CXCL13 + KLRD1; CCL3L1 + CCL4 + CCL4L2 + CXCL13 + LAG3; CCL3L1 + CCL4 + CCL4L2 + CXCL13 + LAG3 + KLRD1; CCL3L1 + CCL4 + CCL4L2 + LAG3 + KLRD1; CCL3L1 + CCL4 + CXCL13; CCL3L1 + CCL4 + CXCL13 + KLRD1; CCL3L1 + CCL4 + CXCL13 + LAG3; CCL3L1 + CCL4 + CXCL13 + LAG3 + KLRD1; CCL3L1 + CCL4 + KLRD1; CCL3L1 + CCL4 + LAG3 + KLRD1; CCL3L1 + CCL4L2 + CCL3; CCL3L1 + CCL4L2 + CCL3 + CXCL13; CCL3L1 + CCL4L2 + CCL3 + CXCL13 + KLRD1; CCL3L1 + CCL4L2 + CCL3 + CXCL13 + LAG3; CCL3L1 + CCL4L2 + CCL3 + CXCL13 + LAG3 + KLRD1; CCL3L1 + CCL4L2 + CCL3 + KLRD1; CCL3L1 + CCL4L2 + CCL3 + LAG3 + KLRD1; CCL3L1 + CCL4L2 + CXCL13; CCL3L1 + CCL4L2 + CXCL13 + KLRD1; CCL3L1 + CCL4L2 + CXCL13 + LAG3; CCL3L1 + CCL4L2 + CXCL13 + LAG3 + KLRD1; CCL3L1 + CCL4L2 + KLRD1; CCL3L1 + CCL4L2 + LAG3; CCL3L1 + CCL4L2 + LAG3 + KLRD1; CCL3L1 + CXCL13; CCL3L1 + CXCL13 + KLRD1; CCL3L1 + CXCL13 + LAG3; CCL3L1 + CXCL13 + LAG3 + KLRD1; CCL3L1 + LAG3 + KLRD1; CCL3 + CXCL13; CCL3 + CXCL13 + KLRD1; CCL3 + CXCL13

+ LAG3; CCL3 + CXCL13 + LAG3 + KLRD1; CCL3 + LAG3 + KLRD1; CCL4 + CCL3; CCL4 + CCL3 + CXCL13; CCL4 + CCL3 + CXCL13 + KLRD1; CCL4 + CCL3 + CXCL13 + LAG3; CCL4 + CCL3 + CXCL13 + LAG3 + KLRD1; CCL4 + CCL3 + KLRD1; CCL4 + CCL3 + LAG3 + KLRD1; CCL4 + CCL4L2 + CCL3 + CXCL13; CCL4 + CCL4L2 + CCL3 + CXCL13 + KLRD1; CCL4 + CCL4L2 + CCL3 + CXCL13 + LAG3; CCL4 + CCL4L2 + CCL3 + CXCL13 + LAG3 + KLRD1; CCL4 + CCL4L2 + CCL3 + KLRD1; CCL4 + CCL4L2 + CCL3 + LAG3; CCL4 + CCL4L2 + CXCL13; CCL4 + CCL4L2 + CXCL13 + KLRD1; CCL4 + CCL4L2 + CXCL13 + LAG3; CCL4 + CCL4L2 + CXCL13 + LAG3 + KLRD1; CCL4 + CCL4L2 + KLRD1; CCL4 + CCL4L2 + LAG3 + KLRD1; CCL4 + CXCL13; CCL4 + CXCL13 + KLRD1; CCL4 + CXCL13 + LAG3; CCL4 + CXCL13 + LAG3 + KLRD1; CCL4 + LAG3 + KLRD1; CCL4L2 + CCL3; CCL4L2 + CCL3 + CXCL13; CCL4L2 + CCL3 + CXCL13 + KLRD1; CCL4L2 + CCL3 + CXCL13 + LAG3; CCL4L2 + CCL3 + CXCL13 + LAG3 + KLRD1; CCL4L2 + CCL3 + KLRD1; CCL4L2 + CCL3 + LAG3 + KLRD1; CCL4L2 + CXCL13; CCL4L2 + CXCL13 + KLRD1; CCL4L2 + CXCL13 + LAG3; CCL4L2 + CXCL13 + LAG3 + KLRD1; CCL4L2 + LAG3 + KLRD1; CXCL13 + LAG3; CXCL13 + LAG3 + KLRD1; KLRD1; KLRD1 + CCL3; KLRD1 + CCL3L1; KLRD1 + CCL4L2; KLRD1 + CXCL13; KLRD1 + LAG3; all biomarkers of Table 1; all biomarkers of Table 5; or all biomarkers of Table 6.

7. The method of any one of claims 1 to 6, wherein said T-cell activating antigen is a cancer antigen, and wherein preferably said sample is a tumor sample.
8. The method of claim 7, wherein said cancer is a brain metastasis of a non-brain primary tumor, is lung cancer, or is glioblastoma, preferably is a brain metastasis of a non-brain primary tumor or is lung cancer.
9. A method of identifying a TCR binding to a T-cell activating antigen presented on a cell, preferably a cancer cell, of a subject, said method comprising
 - (A) identifying a reactive T-cell according to the method of any one of claims 1 to 4,
 - (B) providing the amino acid sequences of at least the complementarity determining regions (CDRs) of the TCR of the reactive T-cell identified in step (A); and, hereby,
 - (C) identifying a TCR binding to an activating antigen presented on a cell.

10. The method of any one of claims 1 to 9, wherein expression of at least one biomarker of step a) and/or the nucleic acid sequences encoding the amino acid sequences of step (B) is/are determined by single-cell sequencing, preferably by single-cell RNA sequencing.
11. The method of claim 9 or 10, wherein said method comprises further step B1) expressing a TCR comprising at least the CDRs determined in step B) in a host cell, preferably a T-cell.
12. The method of claim 11, wherein said method further comprises further step B2) determining binding of the TCR expressed in step B1) to a T-cell activating antigen, preferably complexed in a major histocompatibility complex (MHC), preferably MHC class I, molecule, preferably in a tetramer assay.
13. The method of claim 11 or 12, wherein said method further comprises step B3) determining recognition of cells presenting a T-cell activating antigen by the TCR expressed in step B1).
14. The method of any one of claims 9 to 13, wherein said method further comprises step B4) producing a soluble TCR comprising at least the CDRs determined in step B) and determining binding of said soluble TCR to a cancer cell and/or to a cancer antigen complexed in a major histocompatibility complex (MHC), preferably MHC class I, molecule.
15. A method of providing a T-cell recognizing a cell presenting a T-cell activating antigen, preferably a cancer cell, said method comprising
 - (i) identifying a TCR binding to a cell presenting a T-cell activating antigen according to the method according to any one of claims 9 to 12,
 - (ii) expressing a TCR comprising at least the complementarity determining regions (CDRs) of the TCR of step (I) in a T-cell, and, thereby,
 - (iii) providing a T-cell recognizing a cell presenting a T-cell activating antigen, preferably a cancer cell.

16. A reactive T-cell identified by the method according to any one of claims 1 to 8 and/or obtained or obtainable by the method according to any one of claims 9 to 14, preferably comprising a T-cell receptor comprising an amino acid sequence of SEQ ID NO:1 and/or SEQ ID NO:2, for use in medicine or for use in treating and/or preventing cancer in a subject.
17. A method of identifying at least one biomarker of reactive T-cells, comprising
 - (I) providing expression data of a plurality of biomarkers of T-cells in a sample of a subject,
 - (II) providing a clustering said plurality of T-cells based on the expression of the biomarkers of step (I);
 - (III) providing amino acid sequences of at least the complementarity determining regions (CDRs) of TCRs of T-cells of step (II);
 - (IV) determining reactivity of T-cells expressing a TCR comprising the CDRs of step (III) to cells presenting a T-cell activating antigen;
 - (V) repeating steps (III) and (IV) at least once for further T-cells clustering with T-cells whose TCRs are determined to be reactive to cells presenting a T-cell activating antigen in step (IV), wherein the TCRs of said further T-cells are non-identical to the TCRs of step (IV);
 - (VI) determining at least one cluster of step (II) comprising the highest fraction of T-cells comprising T-cell receptors recognizing cells presenting a T-cell activating antigen; and
 - (VII) determining at least one biomarker expressed by the highest fraction of T-cells in the cluster determined in step (VI), thereby identifying at least one biomarkers of cancer-reactive T-cells.
18. The subject matter of any one of claims 1 to 17, wherein said T-cell(s) is/are CD8⁺ T-cell(s) or CD4⁺ T-cells, preferably are CD8⁺ T-cell(s).
19. The subject matter of any one of claims 9 to 18, wherein said TCR comprises, preferably consists of, a TCR alpha chain and a TCR beta chain.