

SYSTEMS AND METHODS FOR SEQUENCING T CELL RECEPTORS AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

- [0001] This application claims priority to U.S. Provisional Application No. 62/432,525 filed on December 9, 2016, and U.S. Provisional Application No. 62/508,667, filed on May 19, 2017, the contents of each are incorporated by reference herein, in their entirety and for all purposes.

FIELD

- [0002] The disclosure relates generally to the field of bioinformatics. More particularly, the disclosure relates to systems and methods for sequencing T cell receptors (TCRs), identifying T cell clones in a population of cells and for bulk sequencing. The methods identify high-frequency T cell clones associated with tumor reactivity and patient survival.

BACKGROUND

- [0003] Single or combination therapy with immune checkpoint inhibitors has shown significant therapeutic efficacy in cancer patients. However, the majority of patients either do not respond or only respond transiently, raising fundamental questions about the design of the next generation of immunotherapies. To overcome the immunosuppressive nature of the tumor microenvironment and promote durable responses, dual targeting of coinhibitory and costimulatory pathways inducing a stronger T cell activation, can be performed. In some scenarios, a combination of antibodies might synergistically enhance CD8⁺ T cell effector function, for example by restoring a balance of homeostatic regulators, resulting in tumor rejection and long-term responses. T cell clonal expansion could provide a specific gene signature indicating the molecular mechanism of combination therapy. Existing TCR sequence analysis techniques are unable to accurately and reliably identify TCR sequences and clonal expansion based on short read sequencing data. These and other shortcomings are addressed by the methods and systems described herein.

SUMMARY

- [0004] It is to be understood that both the following general description and the following detailed description are exemplary and explanatory only and are not

restrictive. Provided are methods and systems for sequencing short reads of less than about 100 base pairs of RNA obtained from a T cell (and/or receiving sequence data indicative of the same), aligning the short reads with a reference sequence, wherein the reference sequence does not contain a TCR gene sequence, thereby generating a read set comprising mapped short reads and unmapped short reads, discarding mapped short reads from the read set, assembling the unmapped short reads remaining in the read set into one or more long reads, and generating one or more TCR sequences from the one or more long reads.

[0005] Disclosed are methods for sequencing a T cell receptor (TCR), comprising sequencing short reads of less than about 100 base pairs of RNA obtained from a T cell; aligning the short reads with a reference sequence, wherein the reference sequence does not contain a TCR gene sequence, thereby generating a read set comprising mapped short reads and unmapped short reads; discarding mapped short reads from the read set; assembling the unmapped short reads remaining in the read set into one or more long reads; translating the one or more long reads into corresponding amino acid sequences; fractioning TCR V region and TCR J region amino acid reference sequences into k-strings of about six amino acids, aligning the k-strings with the corresponding amino acid sequences from the translating step, detecting one or more conserved TCR CDR3 residues in the k-strings that map to the corresponding amino acid sequences, scoring the level of conservation detected, and selecting corresponding amino acid sequences with a conservation score above a threshold conservation score, and detecting a candidate CDR3 region amino acid sequence in the selected corresponding amino acid sequences; identifying the nucleic acid sequence of the candidate CDR3 region amino acid sequences in the one or more long reads; aligning the nucleic acid sequence of the one or more long reads upstream of the candidate CDR3 region nucleic acid sequence with one or more TCR V gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR V gene sequence; and aligning the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more TCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR J gene sequence, thereby generating a TCR sequence.

[0006] Disclosed are apparatuses comprising one or more processors; and a memory comprising processor executable instructions that, when executed by the one or more processors, cause the apparatus to receive sequence data comprising short reads of less than about 100 base pairs of RNA obtained from a T cell; align the short reads with a reference sequence, wherein the reference sequence does not contain a TCR gene sequence, thereby generating a read set comprising mapped short reads and unmapped short reads; discard mapped short reads from the read set; assemble the unmapped short reads remaining in the read set into one or more long reads; translate the one or more long reads into corresponding amino acid sequences; fraction TCR V region and TCR J region amino acid reference sequences into k-strings of about six amino acids, aligning the k-strings with the corresponding amino acid sequences from the translating step, detecting one or more conserved TCR CDR3 residues in the k-strings that map to the corresponding amino acid sequences, scoring the level of conservation detected, and selecting corresponding amino acid sequences with a conservation score above a threshold conservation score, and detecting a candidate CDR3 region amino acid sequence in the selected corresponding amino acid sequences; identify the nucleic acid sequence of the candidate CDR3 region amino acid sequences in the one or more long reads; align the nucleic acid sequence of the one or more long reads upstream of the candidate CDR3 region nucleic acid sequence with one or more TCR V gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR V gene sequence; and align the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more TCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR J gene sequence, thereby generating a TCR sequence.

[0007] Disclosed are computer readable media, having computer executable instructions embodied thereon, configured for performing a method comprising sequencing short reads of less than about 100 base pairs of RNA obtained from a T cell; aligning the short reads with a reference sequence, wherein the reference sequence does not contain a TCR gene sequence, thereby generating a read set comprising mapped short reads and unmapped short reads; discarding mapped short reads from the read set; assembling the unmapped short reads remaining in the read

set into one or more long reads; translating the one or more long reads into corresponding amino acid sequences; fractioning TCR V region and TCR J region amino acid reference sequences into k-strings of about six amino acids, aligning the k-strings with the corresponding amino acid sequences from the translating step, detecting one or more conserved TCR CDR3 residues in the k-strings that map to the corresponding amino acid sequences, scoring the level of conservation detected, and selecting corresponding amino acid sequences with a conservation score above a threshold conservation score, and detecting a candidate CDR3 region amino acid sequence in the selected corresponding amino acid sequences; identifying the nucleic acid sequence of the candidate CDR3 region amino acid sequences in the one or more long reads; aligning the nucleic acid sequence of the one or more long reads upstream of the candidate CDR3 region nucleic acid sequence with one or more TCR V gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR V gene sequence; and aligning the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more TCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR J gene sequence, thereby generating a TCR sequence.

[0008] Disclosed are methods for sequencing a BCR, comprising sequencing short reads of less than about 100 base pairs of RNA obtained from a B cell; aligning the short reads with a reference sequence, wherein the reference sequence does not contain a BCR gene sequence, thereby generating a read set comprising mapped short reads and unmapped short reads; discarding mapped short reads from the read set; assembling the unmapped short reads remaining in the read set into one or more long reads; translating the one or more long reads into corresponding amino acid sequences; fractioning BCR V region and BCR J region amino acid reference sequences into k-strings of about six amino acids, aligning the k-strings with the corresponding amino acid sequences from the translating step, detecting one or more conserved BCR CDR3 residues in the k-strings that map to the corresponding amino acid sequences, scoring the level of conservation detected, and selecting corresponding amino acid sequences with a conservation score above a threshold conservation score, and detecting a candidate CDR3 region amino acid sequence in

the selected corresponding amino acid sequences; identifying the nucleic acid sequence of the candidate CDR3 region amino acid sequences in the one or more long reads; aligning the nucleic acid sequence of the one or more long reads upstream of the candidate CDR3 region nucleic acid sequence with one or more BCR V gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate BCR V gene sequence; and aligning the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more BCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate BCR J gene sequence, thereby generating a BCR sequence.

[0009] Disclosed are apparatuses comprising one or more processors; and a memory comprising processor executable instructions that, when executed by the one or more processors, cause the apparatus to receive sequence data comprising short reads of less than about 100 base pairs of RNA obtained from a B cell; align the short reads with a reference sequence, wherein the reference sequence does not contain a BCR gene sequence, thereby generating a read set comprising mapped short reads and unmapped short reads; discard mapped short reads from the read set; assemble the unmapped short reads remaining in the read set into one or more long reads; translate the one or more long reads into corresponding amino acid sequences; fraction BCR V region and BCR J region amino acid reference sequences into k-strings of about six amino acids, aligning the k-strings with the corresponding amino acid sequences from the translating step, detecting one or more conserved BCR CDR3 residues in the k-strings that map to the corresponding amino acid sequences, scoring the level of conservation detected, and selecting corresponding amino acid sequences with a conservation score above a threshold conservation score, and detecting a candidate CDR3 region amino acid sequence in the selected corresponding amino acid sequences; identify the nucleic acid sequence of the candidate CDR3 region amino acid sequences in the one or more long reads; align the nucleic acid sequence of the one or more long reads upstream of the candidate CDR3 region nucleic acid sequence with one or more BCR V gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate BCR V gene sequence; and align the nucleic acid sequence of the one or

more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more TCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate BCR J gene sequence, thereby generating a BCR sequence.

[0010] Disclosed are computer readable media, having computer executable instructions embodied thereon, configured for performing a method comprising sequencing short reads of less than about 100 base pairs of RNA obtained from a B cell; aligning the short reads with a reference sequence, wherein the reference sequence does not contain a BCR gene sequence, thereby generating a read set comprising mapped short reads and unmapped short reads; discarding mapped short reads from the read set; assembling the unmapped short reads remaining in the read set into one or more long reads; translating the one or more long reads into corresponding amino acid sequences; fractioning BCR V region and BCR J region amino acid reference sequences into k-strings of about six amino acids, aligning the k-strings with the corresponding amino acid sequences from the translating step, detecting one or more conserved BCR CDR3 residues in the k-strings that map to the corresponding amino acid sequences, scoring the level of conservation detected, and selecting corresponding amino acid sequences with a conservation score above a threshold conservation score, and detecting a candidate CDR3 region amino acid sequence in the selected corresponding amino acid sequences; identifying the nucleic acid sequence of the candidate CDR3 region amino acid sequences in the one or more long reads; aligning the nucleic acid sequence of the one or more long reads upstream of the candidate CDR3 region nucleic acid sequence with one or more BCR V gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate BCR V gene sequence; and aligning the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more BCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate BCR J gene sequence, thereby generating a BCR sequence.

[0011] In some aspects of the disclosed methods, apparatuses, and computer readable media, the short reads are obtained from random-priming of RNA.

[0012] In some aspects of the disclosed methods, apparatuses, and computer readable

media, the T cell is obtained from a human or mouse.

[0013] In some aspects of the disclosed methods, apparatuses, and computer readable media, the reference sequence comprises a human genome, a mouse genome, a human transcriptome, or a mouse transcriptome.

[0014] In some aspects of the disclosed methods, apparatuses, and computer readable media, discarding mapped short reads from the read set further comprises discarding unmapped short reads from the read set that are less than about 35 base pairs or that have a low sequence resolution.

[0015] In some aspects of the disclosed methods, apparatuses, and computer readable media, assembling the unmapped short reads remaining in the read set into one or more long reads comprises aligning the one or more unmapped short reads to one or more TCR sequences from a reference database of TCR sequences; and assembling, based on the alignment, the one or more unmapped short reads into long reads.

[0016] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising appending a TCR C region nucleic acid sequence to the TCR sequence.

[0017] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising, prior to sequencing the short reads of less than about 100 base pairs of RNA obtained from the T cell, administering an immunotherapy to a subject from which the T cell is obtained. In some aspects, the immunotherapy comprises a monotherapy or a combination therapy. In some aspects, the combination therapy comprises a costimulatory agonist and a coinhibitory antagonist.

[0018] In some aspects of the disclosed methods, apparatuses, and computer readable media, repeating all of the steps for a first plurality of T cells of a subject, wherein the T cells are collected prior to administration of a treatment; determining a number of occurrences of unique TCR sequences present in the first plurality of T cells; administering the treatment to the subject; repeating all of the steps for a second plurality of T cells of the subject, wherein the T cells are collected after the administration of the treatment; determining a number of occurrences of unique TCR sequences present in the second plurality of T cells; and determining, based on the numbers of occurrences of unique TCR sequences present in the first plurality of T cells and the second plurality of T cell, one or more unique TCR sequences that experienced clonal expansion.

- [0019] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising determining a T cell clonal expansion signature based on the one or more unique TCR sequences that experienced clonal expansion.
- [0020] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising querying a database of T cell clonal expansion signatures and corresponding treatment responses using the T cell clonal expansion signature; determining, based on the query, the subject's likelihood of responding to the treatment.
- [0021] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising the steps of determining the subject's response to the treatment; storing the T cell clonal expansion signature in a database; and associating the subject's response to the treatment with the T cell clonal expansion signature in the database.
- [0022] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising determining that the TCR sequence is present in a T cell clone that expands in response to a treatment; producing one or more T cells containing the TCR sequence; administering the one or more T cells to a subject; and administering the treatment to the subject.
- [0023] In some aspects of the disclosed methods, apparatuses, and computer readable media, wherein sequencing short reads of less than about 100 base pairs of RNA obtained from a T cell comprises bulk sequencing of short reads of less than about 100 base pairs of RNA obtained from a plurality of T cells.
- [0024] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising performing the steps of aligning the short reads through the step of aligning the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more TCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR J gene sequence, thereby generating a TCR sequence for each of the plurality of T cells.
- [0025] In some aspects of the disclosed methods, apparatuses, and computer readable media, performing steps of aligning the short reads through the step of aligning the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more TCR J gene reference

sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR J gene sequence, thereby generating a TCR sequence for each of the plurality of T cells comprising performing steps of aligning the short reads through the step of aligning the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more TCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR J gene sequence, thereby generating a TCR sequence comprises classifying at least a portion of one or more of steps of aligning the short reads through the step of aligning the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more TCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR J gene sequence, thereby generating a TCR sequence as a job; and distributing a workload for each job across a plurality of processors in parallel.

[0026] In some aspects of the disclosed methods, apparatuses, and computer readable media, the B cell is obtained from a human or mouse.

[0027] In some aspects of the disclosed methods, apparatuses, and computer readable media, assembling the unmapped short reads remaining in the read set into one or more long reads comprises aligning the one or more unmapped short reads to one or more BCR sequences from a reference database of BCR sequences; and assembling, based on the alignment, the one or more unmapped short reads into long reads.

[0028] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising appending a BCR C region nucleic acid sequence to the BCR sequence.

[0029] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising prior to sequencing the short reads of less than about 100 base pairs of RNA obtained from the B cell, administering an immunotherapy to a subject from which the B cell is obtained. In some aspects, the immunotherapy comprises a monotherapy or a combination therapy. In some aspects, the combination therapy comprises a costimulatory agonist and a coinhibitory antagonist.

[0030] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising repeating all of the steps for a first plurality of B cells of a

subject, wherein the B cells are collected prior to administration of a treatment; determining a number of occurrences of unique BCR sequences present in the first plurality of B cells; administering the treatment to the subject; repeating steps a-i for a second plurality of B cells of the subject, wherein the B cells are collected after the administration of the treatment; determining a number of occurrences of unique BCR sequences present in the second plurality of B cells; and determining, based on the numbers of occurrences of unique BCR sequences present in the first plurality of B cells and the second plurality of B cell, one or more unique BCR sequences that experienced clonal expansion.

- [0031] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising determining a B cell clonal expansion signature based on the one or more unique BCR sequences that experienced clonal expansion.
- [0032] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising querying a database of B cell clonal expansion signatures and corresponding treatment responses using the B cell clonal expansion signature; determining, based on the query, the subject's likelihood of responding to the treatment.
- [0033] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising determining the subject's response to the treatment; storing the B cell clonal expansion signature in a database; and associating the subject's response to the treatment with the B cell clonal expansion signature in the database.
- [0034] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising determining that the BCR sequence is present in a B cell clone that expands in response to a treatment; producing one or more B cells containing the BCR sequence; administering the one or more B cells to a subject; and administering the treatment to the subject.
- [0035] In some aspects of the disclosed methods, apparatuses, and computer readable media, sequencing short reads of less than about 100 base pairs of RNA obtained from a B cell comprises bulk sequencing of short reads of less than about 100 base pairs of RNA obtained from a plurality of B cells.
- [0036] In some aspects, disclosed are methods, apparatuses, and computer readable media, further comprising performing the steps of of aligning the short reads with a reference sequence through aligning the nucleic acid sequence of the one or more

long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more BCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate BCR J gene sequence, thereby generating a BCR sequence for each of the plurality of B cells.

[0037] In some aspects of the disclosed methods, apparatuses, and computer readable media, performing the steps of aligning the short reads with a reference sequence through aligning the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more BCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate BCR J gene sequence, thereby generating a BCR sequence for each of the plurality of B cells comprising performing the steps of aligning the short reads with a reference sequence through aligning the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more BCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate BCR J gene sequence, thereby generating a BCR sequence comprises classifying at least a portion of one or more of the steps of aligning the short reads with a reference sequence through aligning the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more BCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate BCR J gene sequence, thereby generating a BCR sequence as a job; and distributing a workload for each job across a plurality of processors in parallel.

[0038] Additional advantages will be set forth in part in the description which follows or may be learned by practice. The advantages will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0039] The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate embodiments and together with the description, serve to explain the principles of the methods and systems.

Figure 1 is a flowchart illustrating a method of random sequencing short repeats and

negative selection of unmapped reads to identify TCR CDR3 sequences.

Figure 2 depicts significant clonal expansion in intratumoral CD8⁺ T cells after treatment.

Figure 3 is a bar chart illustrating significant clonal expansion in intratumoral CD8⁺ T cells after aGITR/aPD1 combination treatment compared to monotherapy.

Figure 4 illustrates T cell lineage after clonal expansion observed using single T cell sequencing.

Figure 5 illustrates intratumoral CD8 T cell clonal analysis based on single cell-sorted RNAseq data.

Figures 6A-6D illustrate T cell activation assays for tumor derived TCRs.

Figures 7A-B illustrate gene expression analysis of single T cell transcriptome;.

Figure 8 illustrates an exemplary operating environment.

Figures 9A-9E show combination therapy expands intratumoral high frequency CD8⁺ T cell clones. **9a**, Schematic of tumor immunotherapy and single cell sorting study design. **9b**, Intratumoral CD8⁺ T cell clonal analysis based on single cell-sorted RNAseq data on day 8 and 11 post tumor challenge. Each circle represents a single CD8⁺ T cell. T cell sharing the same TCR sequence is color-coded, number indicates frequency of individual clone followed by the sequence of CDR3 region of TCR β chain. Day 8: Isotype - SEQ ID NO:290; Anti-GITR – SEQ ID NO:269; Anti-PD1 – SEQ ID NO:270; Day 11: Isotype – SEQ ID NO:291, SEQ ID NO:292 (top to bottom, respectively); Anti-GITR – SEQ ID NO:271-273 (top to bottom, respectively); Anti-PD1 – SEQ ID NO:274-279 (top to bottom, respectively); Anti-GITR +Anti-PD1 – SEQ ID NO:280-289 (top to bottom, respectively) **9c**,

Quantitative analysis of T cell clonality. Data depicts cumulative frequency of expanded CD8⁺ T cell clones from each group (*, $p < 0.05$, **, $p < 0.01$, ***, $p < 0.001$, Fisher's test). **9d**, **9e**, Identification, expression and functional validation of TCRs from intratumoral CD8⁺ T cell clones. High frequency TCR clones were cloned into J.RT3-T3.5 Jurkat T cell line with an AP-1 luciferase reporter gene (**9d**) and their reactivity was tested *in vitro* against MC38 (tumor specific) versus irrelevant tumor cells (B16F10.9 or TRAMP-C2). **9e**, Representative AP-1 luciferase reporter bioassay result with MC38-specific CD8⁺ T cell clone isolated from combination treatment is shown, anti-CD3 and anti-CD28 stimulation was used as positive control.

Figures 10A-D show the identification of unique gene signature in clonally expanded

CD8⁺ T cells. **10a**, Genes upregulated in clonally expanded CD8⁺ T cells from combination therapy comparing to CD8⁺ T cells with isotype treatment or non-expanded CD8⁺ T cells with combination treatment (day 11). Venn diagram shows the number of genes significantly increased ($p < 0.01$, fold change ≥ 2) comparing indicated CD8⁺ T cell population. Heat map shows thirty genes overlapping between the comparison. **10b**, Identification of genes specifically upregulated in clonally expanded CD8 T cells from combination therapy. Schematic and Venn diagram shows different comparison. **10c**, Cumulative distribution function (CDF) plots shows the expression of a uniquely regulated gene by combination treatment, CD226, in total, clonal expanded or non-expanded CD8⁺ T cells. **10d**, Fold changes of CD226 expression level between indicated CD8⁺ T cell population (Number indicates fold changes; NS, not significant).

Figures 11A-11D show combination treatment synergistically regulates CD226/TIGIT pathway favoring anti-tumor immunity. **11a**, FACS analysis of CD226 expression (MFI) on spleen and tumor antigen-specific CD8⁺ T cell populations, gating strategy indicated in FACS plots. Data show one representative experiment out of two independent experiments ($n = 9-10$ mice per group). In the graphs, total CD8 are four bars on left, Non-specific CD8 are four bars in the middle and Ova-specific CD8 are the four bars on the right. **11b**, Schematic shows large unilamellar vesicles (LUVs), reconstituted different components (CD3 and CD226) involved in T cell signaling on the liposomes together with key component of cytosolic tyrosine kinase Ick, Zap70, and SLP76, PI3K (p85a), known to be recruited by phosphorylated costimulatory receptor. **11c**, CD226, but not CD3 ζ , is sensitive to PD-1 bound Shp2. Shp2-containing reactions with increasing PD-1 concentrations terminated at 30min, and subject to SDS-PAGE and phosphotyrosine Western blots (WB). **11d**, FACS analysis of TIGIT expression on spleen and tumor T cell subsets (8 days after tumor implantation). Data show one representative experiment out of two independent experiments ($n = 9-10$ mice per group). (*, $p < 0.05$, **, $p < 0.01$, ***, $p < 0.001$, One-way ANOVA, Tukey's multiple comparison test).

Figures 12A-12H shows CD226 signaling pathway plays a crucial role in mediating anti-tumor response induced by combination treatment. **12a**, MC38 tumor bearing mice were treated with either CD226 blocking Ab or isotype IgG prior to immunotherapy with anti-GITR + anti-PD-1 or isotype IgGs. Percentage of survival

are shown here. Data show one representative experiment out of three independent experiments ($n = 5$ mice per group). **12b**, CD226 KO mice or WT littermates were challenged with MC38 tumor cells and treated with either anti-GITR + anti-PD-1 Ab or isotype Abs. on day 6, 13 after tumor implantation. Data show one representative experiment out of two independent experiments ($n=7-8$ mice per group). (**12c-12e**) Effectiveness of combination treatment doesn't rely on CD28, OX40 and 4-1BB pathway. **12c**, Blocking CD28 signaling with CTLA-4-Ig (10 mg/kg). **12d**, Blocking OX40 signaling with OX40L blocking antibody (10 mg/kg). **12e**, Blocking 4-1BB signaling with 4-1BBL blocking antibody (10 mg/kg). Data shown are survival curves ($n=10$ mice per group). **12f**, Schematic of anti-PD1 clinical study. **12g**, RNA-seq analysis of tumor biopsies shows that *CD226* transcripts were significantly increased after anti-PD-1 Ab treatment in advanced cancer patients ($n=43$, paired *t*-test). **12h**, TCGA data analysis of *CD226* transcript expression level and overall survival in patients with indicated cancer types. (*, $p < 0.05$, **, $p < 0.01$, ***, $p < 0.001$, Log-rank test for survival analysis).

Figure 13 is a table showing TCR-negative cancer or non-cancer cell lines were used as negative controls for MiTCR, TCRklass and rpsTCR pipeline – Negative controls: human and mouse non-T cell lines (2x100bp).

Figure 14 is a table showing the Comparison of TCR detection rate between TCR targeting sequencing and rpsTCR pipeline in single cell sequencing.

Figure 15 is a table showing the pathways specifically upregulated in clonal expanded $CD8^+$ T cells with antibody treatment on day 11.

Figures 16A-16J show that anti-GITR + anti-PD-1 combination synergistically rejects established tumors and reinvigorates dysfunctional T cells. **16a**, MC38 tumor growth in wild-type C57BL/6 mice with anti-GITR and/or anti-PD-1 Ab treatment (day 6, 13). Results depict tumor growth curves of individual mice (cumulative data from two experiments, $n = 17$ mice per group). Numbers on the top represent tumor-free mice over total number of mice in treatment group. **16b**, Cumulative survival curves with indicated treatment. **16c**, **16d**, $CD8^+$ T cell-dependent long-term tumor protection mediated by combination treatment. C57BL/6 mice were treated with anti-CD8, CD4 or anti-CD25 depletion antibody followed by therapy with anti-GITR + anti-PD-1 Ab or control Ab (see Method). **c**. Representative FACS plots showing depletion efficiency by different Abs. **16d**. Average tumor growth curve upon treatment with

different depletion Ab, showing one representative of two experiments ($n = 5$ mice per group). **16e, 16f**, Combination treatment increases intratumoral effector T cell / T_{reg} ratio. **16e**, Representative FACS plots showing tumor T cell subsets on day 11 (FoxP3 versus CD8, cells pre-gated on Live/single cells/CD45⁺/CD3⁺). **16f**, Summary FACS result of intratumoral CD8⁺ T cell / T_{reg} and CD4⁺ T_{Eff} cell / T_{reg} ratio on day 8 and 11 after tumor challenge. Data are representative of three independent experiments ($n = 6-7$ mice per group). **16g-16i**, Combination treatment reinvigorates intratumoral dysfunctional T cells. Tumors were harvested on day 11~12 after implantation, dissociated into single cell suspension restimulated with PMA/Ionomycin with the presence of BFA. Cells were fixed and permeabilized, followed by intracellular staining with Ki67 (**16g**), granzyme A (**16h**), and granzyme B (**16i**). Data shown are percentage of positive cells. ($n=5-10$ mice per group). **16j**, Anti-GITR + anti-PD-1 synergistically rejects established mouse RENCA tumors. Balb/c mice were challenged with 1×10^6 RENCA tumor cells subcutaneously on day 0 and treated with anti-GITR (5 mg/kg) and or anti-PD-1 (5 mg/kg) Ab treatment on day 6, 13, and 20 after tumor implantation. Data shown are percentage of survival ($n = 10$ mice per group). (*, $p < 0.05$, **, $p < 0.01$, ***, $p < 0.001$, ****, $p < 0.0001$; **16b**, **16j** Log-rank test; **16f**, **16g**, One-way ANOVA, Tukey's multiple comparison test). Figure 17 shows an overview of rpsTCR pipeline. Schematic of rpsTCR pipeline, a bioinformatic pipeline for TCR repertoire analysis using random priming short RNA-seq data.

Figure 18 shows rpsTCR pipeline platform validation using human and mouse primary blood cells. Targeted TCR-seq data from healthy human PBMC samples or mouse whole blood were used as a positive control to evaluate false positive or false negative rates comparing to TCRklass alone. Majority of unique CDR3 sequences were detected by the rpsTCR pipeline, MiTCR, and TCRklass, as indicated by the number in Venn diagram. The squared correlations (R^2) between the rpsTCR pipeline, MiTCR, and TCRklass were indicated in the figure.

Figure 19 shows TCRs from LCMV-specific T cell clones (P14) were used as positive control for the bioassay platform. P14 TCR engineered reporter cell line show specificity to LCMV infected cells but not to tumor cell lines. Anti-CD3 and CD28 stimulation was used as positive control for TCR complex signaling.

Figure 20 shows a violin plot showing expression level of thirty genes significantly

enriched in clonally expanded CD8⁺ T cells from combination treatment. Data shown are gene expression (Log2 RPMK) in individual sequenced CD8⁺ T cells (cExp, combination treatment, clonal expanded T cell; cNon, combination treatment, non-expanded; Isotype, total CD8 from isotype treatment group).

Figures 21A-21C show combination treatment expands tumor-antigen specific CD8⁺ T cells with effector function. **a**, Validation of surface expression of OVA peptide-Kb complex on MC38-OVA- β_2m -K^b cells by FACS. MC38-OVA- β_2m -K^b or empty vector control MC38 cells are stained with isotype or anti-Kb-SIINFEKL Ab.

Representative histogram is shown. **b**, Frequency (left) and counts normalized to tumor weight (count/mg, right) of tumor and spleen OVA-specific CD8⁺ T cells from MC38-OVA- β_2m -K^b bearing mice treated with anti-GITR and/or anti-PD-1 Ab.

(Representative of two experiments, $n = 9-10$ mice per group). **c**, Increase of OVA-specific recall response in spleen and tumor CD8⁺ T cells with combination treatment.

Cells were restimulated with or without SIINFEKL peptide in the presence of BFA.

Intracellular expression of IFN γ was analyzed by FACS. Data are representative of two experiments, $n = 9-10$ mice per group, (*, $p < 0.05$, **, $p < 0.01$; ***, $p < 0.001$,

b, One-way ANOVA, **c**, Two-way ANOVA, Tukey's multiple comparison test).

Figures 22A-22D show TIGIT expression at single cell RNA level and FACS analysis of TIGIT/CD226 expression level on different T cell subsets. Cumulative distribution function (CDF) plots showing TIGIT RNA expression (**a**) and FACS analysis (**b**),

representative of two experiments, $n = 9-10$ mice per group) showing expression of TIGIT in total, clonally expanded (OVA-specific) or non-expanded (OVA-non-specific) CD8⁺ T cells.

c, FACS analysis of TIGIT expression on spleen and tumor T cell subsets 8 days after MC38 wild type tumor implantation, percentage of TIGIT⁺ cells within each population are shown.

d, FACS analysis of CD226 expression on spleen and tumor T cell subsets 11 days after MC38 wild type tumor implantation, percentage of CD226⁺ cells within each population are shown.

Data are representative of two experiments, $n = 9-10$ mice per group (*, $p < 0.05$, **, $p < 0.01$, ***, $p < 0.001$,

One-way ANOVA, Tukey's multiple comparison test).

Figures 23A-23F show CD226^{-/-} mice show normal T cell development and

homeostatic function. **a**, Targeting strategy. Coding exons 1 to 2 of mouse *CD226*

was replaced with self-deleting eGFP-Neo cassette (eGFP-polyA-hUb-EM7-neo-

polyA-Prm-Crei-polyA), beginning just 3' to the start ATG in coding exon 1 to 13 bp

before the 3' end of coding exon 2. The intron between coding exons 1-2 is also deleted. After cassette deletion, eGFP, poly A LoxP and cloning sites (1141 bp) remain. **b**, FACS validation of CD226 deletion on T cell subsets. **c**, FACS analysis of T cell development in thymus (Tconv, conventional T cells; DP, CD4/CD8 double positive; SP, single positive; DN, CD4/CD8 double negative). **d**, T cell subsets in spleen and blood analyzed by FACS. **e**, Inflammatory cytokine secretion upon TCR stimulation. Splenocytes from CD226^{-/-} or wild type (WT) mice were stimulated ex vivo with anti-CD3 + anti-CD28 Ab for 16 hours. Supernatant was collected for indicated cytokine release. **f**, Expression level of PD1 and GITR on spleen and blood T cell subsets from CD226^{-/-} or WT mice. Data shown is Mean fluorescence intensity (MFI) (n=3 mice per group).

Figures 24A-24G show CD226 signaling pathway was required for enhanced tumor surveillance in TIGIT^{-/-} mice. Wild type or TIGIT^{-/-} mice were challenged with MC38 tumors, pre-treated with anti-CD226 or control IgG and either received isotype control (**a**) or anti-GITR+anti-PD-1 combination therapy (**b**). Data shown are average tumor growth curves representative of two experiments (n = 4-5 mice per group). **c**, FACS validate over-expression of CD155 on engineered MC38 tumor cells. **d**, Overexpressing CD155, a ligand for CD226/TIGIT, on MC38 tumor cells slows down tumor growth and synergizes with anti-GITR or anti-PD-1 Ab monotherapy (on day 3, 6, 10 and 13) in promoting tumor rejection and long-term survival. Data shown are average tumor growth curve (n=10 mice per group). **e**, Maintained over-expression of CD155 on engineered MC38 tumor cells in vivo. **f**, Overexpression of CD155 on MC38 tumor cells reduces free CD226 receptor on intratumoral, but not splenic, T cell subsets detected by FACS. **g**, Mice were challenged with MC38 control or CD155 overexpression cells. On day 9 after tumor challenge, tumor and spleen cells were analyzed by FACS for T cell activation. For IFN γ expression, cells were restimulated with PMA/ionomycin prior to intracellular staining (n=10 mice per group). (**, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$, One-way ANOVA, Tukey's multiple comparison test).

Figure 25 shows the combination of local radiation with anti-GITR + anti-PD-1 Ab therapy shows efficacy against large established MC38 tumors. Mice bearing large established MC38 tumors were treated with 100 μ g anti-GITR and/or anti-PD-1 Ab or isotype Abs (day 17, 20, 24 and 27) in combination with 0 or 8 Gy local radiation

(day 18). **a**, Average tumor growth curve. **b**, Individual mouse tumor growth curves. Number of tumor-free mice is indicated in the top left panel. **c**, Survival curves from **b**. Data are representative of two experiments ($n = 5-6$ mice per group). (*, $p < 0.05$, **, $p < 0.01$, Log-rank test).

Figure 26 is a list of primers used in library preparation for TCR α/β repertoire sequencing.

Figure 27 shows a comparison of three methods using positive and negative control data sets.

Figure 28 shows the top CDR3s found by miTCR, TCRklass, and rpsTCR pipeline using 50bp priming reads.

Figure 29 shows the top CDR3s found in full VI-Next dataset using 50bp priming reads.

Figure 30 shows the top CDR3s found by miTCR, TCRklass, and rpsTCR pipeline using 100bp priming reads.

Figure 31 shows the top CDR3s found in full VI-Next dataset using 100 bp priming reads.

Figure 32 shows detection of CDR3 in both single cell sequencing and bulk RNA sequencing.

DETAILED DESCRIPTION

[0040] Before the present methods and systems are disclosed and described, it is to be understood that the methods and systems are not limited to specific methods, specific components, or to particular implementations. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting.

[0041] As used in the specification and the appended claims, the singular forms “a,” “an,” and “the” include plural referents unless the context clearly dictates otherwise. Ranges may be expressed herein as from “about” one particular value, and/or to “about” another particular value. When such a range is expressed, another embodiment includes from the one particular value and/or to the other particular value. Similarly, when values are expressed as approximations, by use of the antecedent “about,” it will be understood that the particular value forms another embodiment. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint, and independently of the other

endpoint.

[0042] “Optional” or “optionally” means that the subsequently described event or circumstance may or may not occur, and that the description includes instances where said event or circumstance occurs and instances where it does not.

[0043] Throughout the description and claims of this specification, the word “comprise” and variations of the word, such as “comprising” and “comprises,” means “including but not limited to,” and is not intended to exclude, for example, other components, integers or steps. “Exemplary” means “an example of” and is not intended to convey an indication of a preferred or ideal embodiment. “Such as” is not used in a restrictive sense, but for explanatory purposes.

[0044] It has been observed in accordance with the disclosure that single or combination therapy with immune checkpoint inhibitors has shown significant therapeutic efficacy in cancer patients. However, the majority of patients either do not respond or only respond transiently, raising fundamental questions about the design of the next generation of immunotherapies. To overcome the immunosuppressive nature of the tumor microenvironment and promote durable responses, dual targeting of coinhibitory and costimulatory pathways inducing a stronger T cell activation, can be performed. In some scenarios, a combination of antibodies might synergistically enhance CD8⁺ T cell effector function, for example by restoring a balance of homeostatic regulators, resulting in tumor rejection and long-term responses. Accurate measurement of clonal expansion as a result of treatment can provide a signature indicative of a subject’s response to single or combination therapy. In one aspect, disclosed herein are methods and systems that can generate one or more TCR sequences from short reads obtained from sequencing one or more T cells of a subject. The methods and systems can determine clonal expansion based on the generation of the one or more TCR sequences to provide a signature indicative of subject response and/or potential response.

[0045] Disclosed are components that can be used to perform the disclosed methods and systems, also referred to as the “rpsTCR” pipeline. These and other components are disclosed herein, and it is understood that when combinations, subsets, interactions, groups, etc. of these components are disclosed that while specific reference of each various individual and collective combinations and permutation of these may not be explicitly disclosed, each is specifically contemplated and described

herein, for all methods and systems. This applies to all aspects of this application including, but not limited to, steps in disclosed methods. Thus, if there are a variety of additional steps that can be performed it is understood that each of these additional steps can be performed with any specific embodiment or combination of embodiments of the disclosed methods.

[0046] The present methods and systems may be understood more readily by reference to the following detailed description of preferred embodiments and the examples included therein and to the Figures and their previous and following description.

[0047] As will be appreciated by one skilled in the art, the methods and systems may take the form of an entirely hardware embodiment, an entirely software embodiment, or an embodiment combining software and hardware aspects. Furthermore, the methods and systems may take the form of a computer program product on a computer-readable storage medium having computer-readable program instructions (e.g., computer software) embodied in the storage medium. More particularly, the present methods and systems may take the form of web-implemented computer software. Any suitable computer-readable storage medium may be utilized including hard disks, CD-ROMs, optical storage devices, or magnetic storage devices.

[0048] Embodiments of the methods and systems are described below with reference to block diagrams and flowchart illustrations of methods, systems, apparatuses and computer program products. It will be understood that each block of the block diagrams and flowchart illustrations, and combinations of blocks in the block diagrams and flowchart illustrations, respectively, can be implemented by computer program instructions. These computer program instructions may be loaded onto a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions which execute on the computer or other programmable data processing apparatus create a means for implementing the functions specified in the flowchart block or blocks.

[0049] These computer program instructions may also be stored in a computer-readable memory that can direct a computer or other programmable data processing apparatus to function in a particular manner, such that the instructions stored in the computer-readable memory produce an article of manufacture including computer-readable instructions for implementing the function specified in the flowchart block or

blocks. The computer program instructions may also be loaded onto a computer or other programmable data processing apparatus to cause a series of operational steps to be performed on the computer or other programmable apparatus to produce a computer-implemented process such that the instructions that execute on the computer or other programmable apparatus provide steps for implementing the functions specified in the flowchart block or blocks.

[0050] Accordingly, blocks of the block diagrams and flowchart illustrations support combinations of means for performing the specified functions, combinations of steps for performing the specified functions and program instruction means for performing the specified functions. It will also be understood that each block of the block diagrams and flowchart illustrations, and combinations of blocks in the block diagrams and flowchart illustrations, can be implemented by special purpose hardware-based computer systems that perform the specified functions or steps, or combinations of special purpose hardware and computer instructions.

[0051] Note that in various instances this detailed disclosure may refer to a given entity performing some action. It should be understood that this language may in some cases mean that a system (e.g., a computer) owned and/or controlled by the given entity is actually performing the action.

[0052] In an aspect, illustrated in **FIG. 1**, disclosed is an rpsTCR method **100** for sequencing a T cell receptor (TCR). The steps of the method **100** can be performed in any order or simultaneously. The method **100** can comprise sequencing a nucleic acid sample to generate sequence data and/or receiving sequence data at **110**. Sequencing the nucleic acid sample can comprise sequencing short reads of less than about 100 base pairs of RNA obtained from a T cell of a subject (e.g., single-end 75 base pair reads). The T cell can be obtained from a human or mouse. The T cell can be acquired and/or sequenced prior to, or after, administration of one or more treatments to the subject. The short reads can be obtained from random-priming of RNA. The random primers can be 4-40 nucleotides in length. In some instances, the random primers can be 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotides in length. Sequencing the nucleic acid sample can generate sequence data that can be stored in a data structure. The data structure can comprise one or more nucleic acid sequences and/or a sample identifier.

[0053] In some embodiments, the sequence data can be obtained or received through

any method. For example, the sequence data can be obtained directly, by performing a sequencing process on a sample. Alternatively, or additionally, the sequence data can be obtained indirectly, for example, from a third party, a database and/or a publication. In some embodiments, the sequence data are received at a computer system, for example, from a data storage device or from a separate computer system.

[0054] In some embodiments, the sequence data can comprise bulk sequence data. The term “bulk sequencing” or “next generation sequencing” or “massively parallel sequencing” refers to any high throughput sequencing technology that parallelizes the DNA and/or RNA sequencing process. For example, bulk sequencing methods are typically capable of producing more than one million polynucleic acid amplicons in a single assay. The terms “bulk sequencing,” “massively parallel sequencing,” and “next generation sequencing” refer only to general methods, not necessarily to the acquisition of greater than 1 million sequence tags in a single run. Any bulk sequencing method can be implemented in the disclosed methods and systems, such as reversible terminator chemistry (e.g., Illumina), pyrosequencing using polony emulsion droplets (e.g., Roche), ion semiconductor sequencing (IonTorrent), single molecule sequencing (e.g., Pacific Biosciences), massively parallel signature sequencing, etc.

[0055] In some embodiments, the sequence data can comprise a plurality of sequencing reads. In some embodiments, the sequencing reads have an average read length of no more than 35, 36, 37, 38, 39, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900 or 1000 nucleotides. In some embodiments, the sequencing reads have an average read length of at least 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200 or 250 nucleotides. In some embodiments the coverage of the sequencing reads is no more than 100x, 90x, 80x, 70x, 60x, 50x, 40x, 30x or 20x. In some embodiments the coverage of the sequencing reads is at least 50x, 45x, 40x, 35x, 30x, 25x, 20x, 19x, 18x, 17x, 16x, 15x, 14x, 13x, 12x, 11x or 10x.

[0056] In some embodiments, the sequence data can be produced by any sequencing method known in the art. For example, in some embodiments the sequencing data are produced using chain termination sequencing, sequencing by ligation, sequencing by synthesis, pyrosequencing, ion semiconductor sequencing, single-molecule real-time sequencing, tag-based sequencing, dilute-`n`-go sequencing, and/or 454 sequencing.

[0057] In some embodiments, the sequence data are the result of a process whereby a nucleic acid amplification process is performed to amplify at least part of one or more genomic locus or transcript, followed by the sequencing of the resulting amplification product. Examples of nucleic acid amplification processes useful in the performance of methods disclosed herein include, but are not limited to, polymerase chain reaction (PCR), LATE-PCR, ligase chain reaction (LCR), strand displacement amplification (SDA), transcription mediated amplification (TMA), self-sustained sequence replication (3SR), Q β replicase based amplification, nucleic acid sequence-based amplification (NASBA), repair chain reaction (RCR), boomerang DNA amplification (BDA) and/or rolling circle amplification (RCA).

[0058] In some embodiments, the method includes the step of performing a sequencing process on a sample. Any sample can be used, so long as the sample contains DNA and/or RNA capable of encoding a TCR. In some embodiments, the sample is from a perspective organ, cell or tissue donor. In some embodiments, the sample is from a perspective organ, cell or tissue recipient. The source of the sample may be, for example, solid tissue, as from a fresh, frozen and/or preserved organ, tissue sample, biopsy, or aspirate; blood or any blood constituents, serum, blood; bodily fluids such as cerebral spinal fluid, amniotic fluid, peritoneal fluid or interstitial fluid, urine, saliva, stool, tears; or cells from any time in gestation or development of the subject.

[0059] The method **100** can comprise aligning the short reads with a reference sequence at **120**. The reference sequence can comprise a reference dataset of species-specific RNA sequences. The reference dataset can be stored in a data structure. The data structure can comprise one or more nucleic acid sequences and/or a identifiers. The reference sequence does not contain a TCR gene sequence (e.g., excludes reference sequences that correspond to gene loci of an adaptive immune cell receptor). The alignment thereby generates a read set comprised of mapped short reads and unmapped short reads. In an aspect, aligning the short reads with a reference sequence can comprise one or more techniques described in Trapnell C., et al. TopHat: discovering splice junctions with RNA-Seq, Bioinformatics (2009) 25 (9):1105-1111. Mapped short reads can be discarded at **130**. A resulting data structure can be generated that comprises the unmapped short reads. The remaining steps of the method can be performed on the unmapped short reads which are

normally discarded and not subjected to further analysis in the TCR context. Such filtering out of mapped short reads represents a departure from state of the art TCR analysis and results in downstream improvements in both accuracy and precision.

[0060] The method **100** can further comprise performing a quality control process on the one or more unmapped short reads. Performing the quality control process on the one or more unmapped short reads can comprise one or more of removing low quality nucleotides or removing very short reads. Removing very short reads can comprise removing any read less than 35 base pairs long.

[0061] In an aspect, the method **100** can assemble the one or more unmapped short reads into one or more long reads for further processing at **140**. In an aspect, assembling the one or more unmapped short reads into one or more long reads for further processing can comprise aligning the one or more unmapped short reads to one or more TCR sequences from a reference database of TCR sequences and assembling the one or more unmapped short reads into long reads (candidate TCR sequences) based on the reference database of TCR sequences. In another aspect, assembling the one or more unmapped short reads into one or more long reads for further processing can comprise assembling the one or more unmapped reads into long reads (candidate TCR sequences) without the use of a reference database of TCR sequences.

[0062] In an aspect, assembling the one or more unmapped short reads into one or more long reads for further processing can comprise one or more techniques disclosed in Warren, R. L., B. H. Nelson, and R. A. Holt. 2009. Profiling model T-cell metagenomes with short reads. *Bioinformatics* 25: 458–464, incorporated herein by reference in its entirety (the iSSAKE platform). In an aspect, assembling the one or more unmapped short reads into one or more long reads for further processing can comprise aligning the one or more unmapped short reads against known, curated V genes of a desired adaptive immune cell receptor. The one or more unmapped short reads with best forward or reverse-complement alignment to 3' end of the V genes with unmatched nucleotides 3' of the V alignment can be labeled as seeds for de novo assembly. The one or more unmapped short reads fully aligning to receptor V genes or constant regions or possible junctions between J genes and constant regions can be discarded from future assembly.

[0063] Each seed sequence can be used to nucleate an assembly. For example, a

subsequence length (k) can begin at the longest unassembled read length. Then the 3'-most subsequence of length k can be generated (k -mer). If the k -mer matches the 5' end bases of one or more forward or reverse-complement read(s) r , the matching read(s) r can be used to extend the assembly (if overhanging extension nucleotides do not agree across r , a majority rule can be used to build a consensus assembly sequence(s)). If there is no match and k is greater than the minimum subsequence length specified by the user, the matching can be repeated with a new k shorter by one base. If there is no match and k equals a minimum subsequence length specified by a user, assembly is complete. Assembly is complete when all seed sequences and resulting assembly sequences reach maximal extension (e.g., user defined). The steps above can be repeated with new assembly sequences. The result is a read set comprising one or more long reads.

[0064] In another aspect, an alternative approach for assembling the one or more unmapped short reads into one or more long reads for further processing can comprise one or more techniques disclosed in Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*. 2011;29(7):644–52, incorporated herein by reference in its entirety (the Trinity platform). Assembling the one or more unmapped short reads into one or more long reads for further processing can comprise a multi-step approach. The first step can comprise assembling the one or more unmapped short reads into unique sequences of transcripts using a greedy k -mer-based approach for transcript assembly, recovering a single (best) representative for a set of alternative variants that share k -mers (owing to alternative splicing, gene duplication or allelic variation). The k -mer-based approach can comprise constructing a dictionary of k -mer forward and reverse-complement subsequences from all candidate TCR sequences (by way of example, $k=25$). The most frequent k -mer in the dictionary can be selected to seed a contig assembly, excluding k -mers with low complexity or only observed once. The seed can be extended in either direction by finding the highest occurring k -mer with a $k - 1$ overlap with the current assembly and concatenating its overhanging nucleotide to the growing assembly sequence. Once a k -mer has been used for extension, it can be removed from the dictionary. Seed extension can be repeated until the assembly cannot be further extended. Selection of most frequent k -mer and seed extension can be repeated with the next most frequent

k -mer until the dictionary is exhausted.

[0065] The second step of the multi-step approach can comprise clustering related contigs that correspond to portions of alternatively spliced transcripts or otherwise unique portions of paralogous genes. A de Bruijn graph can then be constructed for each cluster of related contigs, each graph reflecting the complexity of overlaps between variants. Contigs can be clustered if there is an overlap of $k-1$ nucleotides between contigs and if there is a minimal number of reads that span the junction across both contigs with a $(k-1)/2$ nucleotide match on each side of the $(k-1)$ -mer junction. Grouping can be repeated until no further contigs can be added to any group. A de Bruijn graph can be constructed for each group using a word size of $k-1$ to represent nodes and k to define the edges connecting the nodes. Each edge of the de Bruijn graph can comprise the number of k -mers in the original read set that support it. Each read can be assigned to the group with which it shares the largest number of k -mers and the regions within each read that contribute k -mers to the group can be determined.

[0066] The third step of the multi-step approach can comprise analyzing the paths taken by reads and read pairings in the context of the corresponding de Bruijn graph and outputting plausible transcript sequences, resolving alternatively spliced isoforms and transcripts derived from paralogous genes. Subsequent iteration between merging nodes and pruning edges can be implemented to identify paths that are supported by reads or read pairs and return these paths as long reads (candidate TCR sequences). Merging nodes can comprise merging consecutive nodes in linear paths in the de Bruijn graph to form nodes that represent longer sequences. Pruning edges can comprise pruning edges that represent minor deviations supported by comparatively few reads that likely correspond to sequencing errors. The third step can further comprise performing plausible path scoring, by identifying those paths in the de Bruijn graph that are supported by actual reads and read pairs, using a dynamic programming procedure that traverses potential paths in the graph while maintaining the reads (and pairs) that support them. Because reads and sequence fragments (paired reads) are typically much longer than k , they can resolve ambiguities and reduce the combinatorial number of paths to a much smaller number of actual transcripts, enumerated as linear sequences. The result is a read set comprising one or more long reads.

[0067] TCR sequence assembly can be carried out using one of several platforms such as, but not limited to, the iSSAKE platform and the Trinity platform. Table 1 shows that the Trinity platform and the iSSAKE platform were effectively equivalent for single cell sequencing as a component of the disclosed methods and systems, but the iSSAKE platform was superior at bulk sequencing. The iSSAKE platform utilizes seed sequences for TCRs which can improve performance of the rpsTCR pipeline with bulk assembly.

Table 1.

CDR3	Single cells detected by iSSAKE	Single cells detected by Trinity	Counts in bulk by iSSAKE	Found in bulk by Trinity
CASSPTGYNSPLYF (SEQ ID NO:293)	4	4	116	1
CASSQVQGS AETLYF (SEQ ID NO:294)	5	5	80	0
CASSGTGGNQDTQYF (SEQ ID NO:295)	1	1	0	0
CASGDAGTGNYAEQFF (SEQ ID NO:296)	1	1	19	1
CASSLRTGYNSPLYF (SEQ ID NO:297)	3	3	41	1
CASRLGGDQNTLYF (SEQ ID NO:298)	3	3	49	1
CASKTGGYEQYF (SEQ ID NO:299)	1	1	37	0
CASSEGDTLYF (SEQ ID NO:300)	1	1	13	0
CASSPGTFNQDTQYF (SEQ ID NO:301)	3	3	16	0
CASASWTGDEQYF (SEQ ID NO:302)	1	1	0	0
CASSLPGSQNTLYF (SEQ ID NO:303)	1	1	0	0
CASSRDWAQDTQYF (SEQ ID NO:304)	2	2	58	0
CASSDNWGAGEQYF (SEQ ID NO:305)	1	1	1	1
CASSSGTASDTQYF (SEQ ID NO:306)	1	1	0	0
CASSQTRDWGYEQYF (SEQ ID NO:307)	1	1	33	0
CTCSGGLGGLEQYF (SEQ ID NO:308)	1	1	5	1
CASSLGTGGIEQYF (SEQ ID NO:309)	1	1	3	1
CASSLSDSNQDTQYF (SEQ ID NO:310)	1	1	0	0
CASSERGGRDQYF (SEQ ID NO:311)	1	1	0	0
CTCSAVREGNSPLYF (SEQ ID NO:312)	1	1	3	1
CASSLTGVSNERLFF (SEQ ID NO:313)	1	1	0	0
CASSRQLNSDYTF (SEQ ID NO:314)	2	2	39	1
CASSLRQGSNTEVFF (SEQ ID NO:315)	1	1	15	1
CASSQNRDISAETLYF (SEQ ID NO:316)	1	1	53	0
CASSWTANTEVFF (SEQ ID NO:317)	1	1	0	0
CASSLRDWGQDTQYF (SEQ ID NO:318)	1	1	22	0
CASSHWGGTTGQLYF (SEQ ID NO:319)	1	1	0	0
CASSYSKGS AETLYF (SEQ ID NO:320)	1	1	0	0

CAVSMINYNVLYF (SEQ ID NO:321)	1	1	0	0
CASSDGQNTLYF (SEQ ID NO:322)	1	0	4	1
CASSQEGPGQLYF (SEQ ID NO:323)	1	1	15	1
CASTGQGYNSPLYF (SEQ ID NO:324)	1	0	0	0

[0068] The method **100** can generate one or more TCR sequences from the one or more long reads at **150**. In an aspect, generating one or more TCR sequences from the one or more long reads can comprise one or more techniques disclosed in Yang, X. et al. TCRklass: a new K-string-based algorithm for human and mouse TCR repertoire characterization. J. Immunol. 194, 446–454 (2015), incorporated herein by reference in its entirety. Generating one or more TCR sequences from the one or more long reads can comprise translating each of the one or more long reads on all six frames, comparing each translation frame to a 3-string profile of a reference variable (V) and joining (J) amino acid sequence, identifying the translation frame with a highest number of matched k-strings, determining a position of a conserved residue in the long read by determining a conserved residue support score (S_{cr}) for each residue in the long read from the translation frame with a highest number of matched k-strings, identifying candidate conserved residues with a highest S_{cr} in V and J gene segments of the long read from the translation frame, and identifying a CDR3 region located between two conserved residues in the V and J gene segments as a TCR sequence. In an aspect, the method **100** can further comprise appending a TCR C region nucleic acid sequence to the TCR sequence.

[0069] In another aspect, generating one or more TCR sequences from the one or more long reads can comprise translating the one or more long reads into corresponding amino acid sequences. TCR V region and TCR J region amino acid reference sequences can be fractioned into k-strings of about six amino acids. The k-strings can be aligned with the corresponding amino acid sequences. One or more conserved TCR CDR3 residues can be detected in the k-strings that map to the corresponding amino acid sequences. A detected level of conservation can be scored and corresponding amino acid sequences with a conservation score above a threshold conservation score can be selected. A candidate CDR3 region amino acid sequence can then be detected in the selected corresponding amino acid sequences.

[0070] The nucleic acid sequence of the candidate CDR3 region amino acid sequences in the one or more long reads can be identified. The nucleic acid sequence

of the one or more long reads upstream of the candidate CDR3 region nucleic acid sequence can be aligned with one or more TCR V gene reference sequences. A degree of alignment can be scored and long reads above a threshold alignment score can be identified as comprising a candidate TCR V gene sequence.

[0071] The nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence can be aligned with one or more TCR J gene reference sequences. A degree of alignment can be scored and long reads above a threshold alignment score can be identified as comprising a candidate TCR J gene sequence, thereby generating a TCR sequence. In an aspect, the method **100** can further comprise appending a TCR C region nucleic acid sequence to the TCR sequence

[0072] The method **100** can further comprise comparing the one or more TCR sequences to a TCR sequence library of known TCR sequences and corresponding treatment responses to one or more treatments, identifying which of the one or more TCR sequences have a match in the TCR sequence library with a high corresponding treatment response, and identifying the one or more treatments to which the subject having the one or more TCR sequences is likely to respond. Once the subject has been identified as having a TCR sequence that is likely to respond to a specific treatment, the subject can be administered the specific treatment.

[0073] The method **100** can further comprise performing the method **100** prior to, and after, administration of a treatment of a subject for a disease to assess clonal expansion. For example, a first plurality of T cells of a subject can be collected prior to administration of a treatment. The first plurality of T cells can be sequenced and the method **100** can be performed. A number of occurrences of unique TCR sequences present can be determined. The treatment can be administered to the subject and a second plurality of T cells of the subject can be collected. The second plurality of T cells can be sequenced and the method **100** can be performed. A number of occurrences of unique TCR sequences present can be determined. The numbers of occurrences between the first plurality of T cells and the second plurality of T cells can then be determined. In some instances, a specific TCR sequence can be determined to have experienced clonal expansion. In other instances, some, all, or none of the TCR sequences that experienced clonal expansion between the first plurality of T cells and the second plurality of T cells are the same TCR sequence. A

result is a T cell clonal expansion signature. The T cell clonal expansion signature can comprise one or more of, a number of T cells that experienced clonal expansion, an identifier of T cells that experienced clonal expansion, an overall quantity of clonal expansion, a quantity of clonal expansion per T cell, combinations thereof, and the like. The subject's response to treatment can be recorded and associated with the T cell clonal expansion signature. The process can be repeated for a plurality of subjects, thereby generating a database of T cell clonal expansion signatures and corresponding treatment responses. The disclosed methods and systems can subsequently compare a T cell clonal expansion signature of a new subject to the database to ascertain a likely response to treatment(s) for the subject.

[0074] In some aspects of the method **100**, the subject can be administered an immunotherapy prior to the collection of T cells for sequencing. The immunotherapy can be a monotherapy or a combination therapy. For example, the immunotherapy can be the combination of a costimulatory agonist and a coinhibitory antagonist. In some aspects, T cell inhibitory receptors or receptors on a tumor cell, including, but not limited to, PD1, PDL1, CTLA4, LAG3 and TIM3, can be targeted during the immunotherapy. Thus, in some aspects, the immunotherapy can comprise an antibody or antigen-binding fragment thereof that specifically binds to one or more of PD1, PDL1, CTLA4, LAG3, and TIM3. As part of an immunotherapy regimen, the subject may be administered an antibody or antigen-binding fragment thereof that specifically binds to one or more of PD1, PDL1, CTLA4, LAG3, and TIM3, or may be administered any combination of two or more such antibodies or antigen-binding fragments thereof.

[0075] In some aspects, the immunotherapy comprises administering to the patient an antibody or antigen-binding fragment thereof that binds to PD1. In some preferred embodiments, the antibody or antigen-binding fragment thereof that binds to PD1 comprises at least the heavy chain variable region (HCVR) sequence of SEQ ID NO:21 and the light chain variable region (LCVR) sequence of SEQ ID NO:22. In aspects, any of the antibodies or antigen-binding fragments thereof that bind PD1 can be any of the antibodies or antigen-binding fragments thereof described in US Application No. 14/603,776 (Publication No. US 2015-0203579), which is hereby incorporated by reference herein. For example, in some embodiments, the antibody or antigen-binding fragment thereof that binds to PD1 comprises a HCVR having an

amino acid sequence from among the sequences listed in Table 2 and a LCVR. In some embodiments, the antibody or antigen-binding fragment thereof that binds to PD1 comprises a LCVR having an amino acid sequence from among the sequences listed in Table 2 and an HCVR. In some embodiments, the antibody or antigen-binding fragment thereof that binds to PD1 comprises an HCVR and LCVR pair as shown in Table 2. Other antibodies that bind to PD1 can be used (or antigen-binding fragments thereof), and these include but are not limited to pembrolizumab, nivolumab, durvalumab, atezolizumab, pidilizumab, camrelizumab, PDR001, MED10680, JNJ-63723283, and MCLA-134.

Table 2: Amino Acid Sequence Identifiers for PD1 antibodies

HCVR SEQ ID NO:	LCVR SEQ ID NO:		HCVR SEQ ID NO:	LCVR SEQ ID NO:
1	2		28	26
3	4		29	26
5	6		30	26
7	8		31	26
9	10		32	26
11	12		33	26
13	14		34	26
15	16		35	26
17	18		36	26
19	20		37	26
21	22		38	24
23	24		39	24
25	26		40	24
27	26			

[0076] In some aspects, the immunotherapy comprises administering to the patient an antibody or antigen-binding fragment thereof that binds to the LAG3 protein (aka CD223). In some aspects, the antibody or antigen-binding fragment thereof that binds to LAG3 comprises at least the HCVR sequence of SEQ ID NO:93 and the LCVR sequence of SEQ ID NO:94. In some aspects, the antibodies or antigen-binding fragments thereof that bind LAG3 can be any of the antibodies or antigen-binding fragments thereof described in US Application No 15/289,032 (Publication No. US 2017-0101472), which is hereby incorporated by reference herein. For example, in some aspects, the antibody or antigen-binding fragment thereof that binds to LAG3 comprises a HCVR having an amino acid sequence from among the sequences listed in Table 3 and a LCVR. In some aspects, the antibody or antigen-binding fragment thereof that binds to LAG3 comprises a LCVR having an amino acid sequence from among the sequences listed in Table 3 and an HCVR. In some embodiments, the antibody or antigen-binding fragment thereof that binds to LAG3 comprises an HCVR and LCVR pair as shown in Table 3. Other antibodies that bind to LAG3 can be used (or antigen-binding fragments thereof), and these include but are not limited to BMS-986016 and GSK2381781.

Table 3: Amino Acid Sequence Identifiers for LAG3 antibodies

HCVR	LCVR		HCVR	LCVR
SEQ ID NO:	SEQ ID NO:		SEQ ID NO:	SEQ ID NO:
41	42		81	82
43	44		83	84
45	46		85	86
47	48		87	88
49	50		89	90
51	52		91	92
53	54		93	94
55	56		95	96
57	58		97	98
59	60		99	98
61	62		100	98
63	64		101	98

65	66		102	98
67	68		103	98
69	70		104	105
71	72		106	105
73	74		107	105
75	76		108	109
77	78		110	111
79	80			

[0077]

[0078] In some aspects, the immunotherapy comprises administering to the patient an antibody or antigen-binding fragment thereof that binds to PDL1. In some preferred aspects, the antibody or antigen-binding fragment thereof that binds to PDL1 comprises at least the HCVR sequence of SEQ ID NO:122 and the LCVR sequence of SEQ ID NO:123. In some aspects, the antibodies or antigen-binding fragments thereof that bind PDL1 can be any of the antibodies or antigen-binding fragments thereof described in US Application No 14/603,808 (Publication No. US 2015-0203580), which is hereby incorporated by reference herein. For example, in some aspects, the antibody or antigen-binding fragment thereof that binds to PDL1 comprises a HCVR having an amino acid sequence from among the sequences listed in Table 4 and a LCVR. In some aspects, the antibody or antigen-binding fragment thereof that binds to PDL1 comprises a LCVR having an amino acid sequence from among the sequences listed in Table 4 and an HCVR. In some aspects, the antibody or antigen-binding fragment thereof that binds to PDL1 comprises an HCVR and LCVR pair as shown in Table 4. Other antibodies that bind to PDL1 can be used (or antigen-binding fragments thereof), and these include but are not limited to, one or more of avelumab, atezolizumab, and durvalumab.

Table 4: Amino Acid Sequence Identifiers for PDL1 antibodies

HCVR	LCVR		HCVR	LCVR
SEQ ID NO:	SEQ ID NO:		SEQ ID NO:	SEQ ID NO:
112	113		137	138
114	115		139	140

116	117		141	142
118	119		143	144
120	121		145	146
122	123		147	146
124	125		148	146
126	127		149	146
128	129		150	146
130	131		151	146
132	133		152	146
134	133		153	146
135	136		154	146

[0079] In some aspects, the immunotherapy comprises administering to the patient an antibody or antigen-binding fragment thereof that binds to CTLA4. In some aspects, the antibodies or antigen-binding fragments thereof that bind CTLA4 can be any of the antibodies or antigen-binding fragments thereof described in US Provisional Application No 62/537,753, filed on July 27, 2017, which is hereby incorporated by reference herein. For example, in some aspects, the antibody or antigen-binding fragment thereof that binds to CTLA4 comprises a HCVR having an amino acid sequence from among the sequences listed in Table 5 and a LCVR. In some aspects, the antibody or antigen-binding fragment thereof that binds to CTLA4 comprises a LCVR having an amino acid sequence from among the sequences listed in Table 5 and an HCVR. In some aspects, the antibody or antigen-binding fragment thereof that binds to CTLA4 comprises an HCVR and LCVR pair as shown in Table 5. Other antibodies that bind to CTLA4 can be used (or antigen-binding fragments thereof), and these include but are not limited to, one or more of ipilimumab and tremelimumab, as well as any of the antibodies or antigen-binding fragments thereof disclosed in US Patent Nos. 6,984,720; 7,605, 238; or 7,034,121, all of which are hereby incorporated by reference herein.

Table 5: Amino Acid Sequence Identifiers for CTLA4 antibodies

HCVR	LCVR	HCVR	LCVR
155	156	187	188
157	158	189	190
159	160	191	192
161	162	193	192
163	164	195	194
165	166	197	196
167	168	199	198
169	170	201	200
171	172	203	202
173	174	205	204
175	176	207	206
177	178	209	208
179	180	211	210
181	182	213	212
183	184	215	214
185	186	217	216

[0080] In some aspects, the immunotherapy comprises administering to the patient an antibody or antigen-binding fragment thereof that binds to GITR. In some preferred embodiments, the antibody or antigen-binding fragment thereof that binds to GITR comprises at least the HCVR sequence of SEQ ID NO:261 and the LCVR sequence of SEQ ID NO:259. In aspects, any of the antibodies or antigen-binding fragments thereof that bind GITR can be any of the antibodies or antigen-binding fragments thereof described in US Application No. 15/619,068, which is hereby incorporated by reference herein. For example, in some aspects, the antibody or antigen-binding fragment thereof that binds to GITR comprises a HCVR having an amino acid sequence from among the sequences listed in Table 6 and a LCVR. In some embodiments, the antibody or antigen-binding fragment thereof that binds to GITR comprises a LCVR having an amino acid sequence from among the sequences listed in Table 6 and an HCVR. In some aspects, the antibody or antigen-binding fragment

thereof that binds to GITR comprises an HCVR and LCVR pair as shown in Table 6. Other known antibodies that bind to GITR can be used (or antigen-binding fragments thereof).

Table 6: Amino Acid Sequence Identifiers for GITR antibodies

HCVR	LCVR		HCVR	LCVR
218	219		248	249
220	221		250	251
222	223		252	253
224	225		254	255
226	227		256	257
228	229		258	259
230	231		260	259
232	233		261	259
234	235		262	259
236	237		263	259
238	239		264	259
240	241		265	259
242	243		266	259
244	245		267	259
246	247		268	259

[0081] In some aspects, a TCR sequence can be identified as a sequence present in a T cell clone that expands in response to a particular treatment. These identified TCR sequences can be used for T cell therapy. For example, the identified TCR sequence can be used to produce T cells containing this particular TCR sequence. These T cells containing the identified TCR sequence can then be administered to a subject who in turn can then be treated with the particular treatment to which the TCR sequence was determined to respond. In some aspects, the T cell therapy can be administered to the same subject from which the TCR sequence was identified in order to increase the number of T cells responding to the particular treatment. In some aspects, the T cell therapy can be administered to a subject other than the one from which the TCR sequence was identified. Administering the T cell therapy to a subject other than the one from which the TCR sequence was identified gives a subject who otherwise would not necessarily have responded to the particular treatment the ability to respond

to the particular treatment.

[0082] In some aspects, TCR signaling can be studied in response to particular drugs for those T cells containing the identified TCR sequences. The TCR signaling of those receptors having a specific TCR sequence present in T cells that expand to particular treatments provides insight into tumor immune surveillance.

[0083] Another use of the identified TCR sequences can be for determining a target for treating a tumor present in the subject with the identified TCR sequences. The antigen that binds the identified TCR sequence is a target for the tumor present in that subject. Once a target has been identified, treatments can then be determined.

[0084] In some aspects, identification of TCR sequences in clonal expansion can be used for prognosis of both viral and bacterial infections and can be used to monitor disease progress of cancer and infectious diseases.

[0085] As shown in **FIGs. 27-33**, the discarding of mapped short reads in step **130** contributes to an improvement in both accuracy and precision as compared to state of the art TCR analysis techniques.

[0086] The methods of **FIGs. 27-33** use negative control datasets expected to have no TCRs. These samples are all random priming RNA-Seq with reads of length 100 base pairs (bp). Five datasets are from various mouse tumor cell lines (these are shown in Table 7). Two datasets are from spleen samples of mice with Rag1/2 knocked out, which is a gene required for the formation of TCRs. Two datasets are from human cell lines, one of neural progenitor cells (NeuProgCell) and the other of myoblasts (LHCN-M2). The positive control dataset (VI-next-mouse-T cell) is a targeted TCR sequencing of a healthy B6 mouse sample with reads of 300bp in length. This dataset was manipulated to create simulated testing datasets of various sequencing depths (10 million, 50 million, 100 million, 200 million, or 500 million reads) and read lengths (50bp or 100bp).

[0087] Another testing dataset is bulk random priming RNA-Seq with read length of 80bp of sorted T cells from mouse tumor samples. The corresponding positive control datasets consist of single-cell RNA-Seq with read length 75bp from the C1 Fluidigm platform of the same sorted T cells as the bulk dataset.

[0088] The data sets for TCR pipeline benchmarks are shown in Table 7.

Table 7.

Classification	File name	Sample type	Sequencing platform	Read length
Negative control	M620270	MC38 colon tumor cell line	Random priming	100bp
Negative control	M620272	MC38 colon tumor cell line	Random priming	100bp
Negative control	M620295	B16F1 melanoma tumor cell line	Random priming	100bp
Negative control	M620279	Colon26 tumor cell line	Random priming	100bp
Negative control	M620343	Renca kidney tumor cell line	Random priming	100bp
Negative control	T-ALL_neg1	Spleen from Rag1/2 KO mouse	Random priming	100bp
Negative control	T-ALL_neg2	Spleen from Rag1/2 KO mouse	Random priming	100bp
Negative control	NeuProgCell	Spleen from Rag1/2 KO mouse	Random priming	100bp
Negative control	LHCN-M2	Human cell line	Random priming	100bp
Positive control	VI-next-mouse-T cell	Healthy B6 mouse T cells	Targeted PCR (fastq)	300bp Merged
Positive control	T-ALL Bulk RNA	T-All sample	Targeted PCR (fasta)	2x300bp
Positive control	corresponding C1 data	MC38/41BBL-Puro tumor T cells	Random priming	75bp
Testing sample	MP-100bp-10M	Healthy B6 mouse T cells	Random priming	100bp
Testing sample	MP-100bp-50M	Healthy B6 mouse T cells	Random priming	100bp
Testing sample	MP-100bp-100M	Healthy B6 mouse T cells	Random priming	100bp
Testing sample	MP-100bp-200M	Healthy B6 mouse T cells	Random priming	100bp
Testing sample	MP-100bp-500M	Healthy B6 mouse T cells	Random priming	100bp
Testing sample	MP-50bp-10M	Healthy B6 mouse T cells	Random priming	50bp
Testing sample	MP-50bp-50M	Healthy B6 mouse T cells	Random priming	50bp
Testing sample	MP-50bp-100M	Healthy B6 mouse T cells	Random priming	50bp
Testing sample	MP-50bp-200M	Healthy B6 mouse T cells	Random priming	50bp
Testing sample	MP-50bp-500M	Healthy B6 mouse T cells	Random priming	50bp
Testing sample	T-ALL	T-ALL	Random priming	100bp
Testing sample	Bulk RNA	MC38/41BBL-Puro tumor T cells	Random priming	80bp

[0089] **FIG. 27** shows that the sensitivity of the rpsTCR pipeline is very similar to TCRklass and better than miTCR. Very few, if any, false positives are detected using the rpsTCR pipeline. The rpsTCR pipeline identifies no TCRs in any of the negative control datasets. In comparison, TCRklass identifies very few and MiTCR identifies on average tens of TCRs in each dataset. The rpsTCR pipeline and TCRklass show comparable sensitivity of around 80% in identifying TCR CDR3s in the VI-next-mouse-T cell positive control dataset, whereas MiTCR has lower sensitivity closer to 75%.

[0090] **FIG. 18** depicts the ability of the rpsTCR pipeline to identify positive data not detected by MiTCR or TCRklass. There is a strong correlation in TCR counts identified in the positive control VI-next dataset amongst all three methods (MiTCR, TCRklass, and the rpsTCR pipeline), with around 10% of total TCRs identified by the rpsTCR pipeline method and TCRklass and not by MiTCR.

- [0091] **FIGs 28 and 30** shows CDR3 detection comparison between MiTCR, TCRklass, and the rpsTCR pipeline using random priming reads of 50bp or 100bp, respectively, of mouse T cells. The sequence of each of the top CDR3s found is listed. **FIGs 29 and 31** shows the top CDR3s found in full VI-next dataset using 50bp or 100bp, respectively. The rpsTCR pipeline shows a higher sensitivity.
- [0092] **FIGs 28 and 29** outline the comparison of CDR3s found in the simulated 50 base pair (bp) read length datasets formed by subsampling the VI-next positive control dataset. Unlike MiTCR and TCRklass, the rpsTCR pipeline implements an assembly of the short 50bp reads into longer contigs to better identify TCR sequences. The top CDR3s found by MiTCR in the dataset of 500 million reads (MP-50bp-500M) were not found by any method in the positive control dataset or in MP-50bp-500M by TCRklass or the rpsTCR pipeline (**FIG. 28**, left table). The top CDR3s found by TCRklass in MP-50bp-500M were found by the other methods in both the test dataset and the VI-next dataset (**FIG. 28**, middle table). The top CDR3s identified by the rpsTCR pipeline were identified in larger numbers by all methods in the VI-next dataset and partially by the other two methods in the MP-50bp-500M dataset (**FIG. 28**, right table). The top CDR3s found in the VI-next dataset were found most often by the rpsTCR pipeline in subsampled MP to 10-500 million reads of 50bp in length (**FIG. 29**).
- [0093] **FIGs 30 and 31** outline the comparison of CDR3s found in the simulated 100bp datasets formed by subsampling 200 million reads from the VI-next positive control dataset to compare the rpsTCR pipeline method without the assembly step to MiTCR and TCRklass. Again, the top CDR3s found by MiTCR in the 500 million reads of length 100bp (MP-100bp-500M) were not found by any other method (**FIG. 30**, left table). The top two CDR3s found by TCRklass appear to be false positives that were not found by any of the other methods (**FIG. 30**, middle table), but the others correspond well to the top CDR3s found by the rpsTCR pipeline (**FIG. 30**, right table). All methods have comparable sensitivity at this read length in identifying the top CDR3s from the VI-next dataset (**FIG. 31**).
- [0094] **FIG. 32** shows that more than half of the CDR3s detected in single cell sequencing also can be detected in bulk RNA sequencing. The final comparison of the three methods involves quantifying how many of the TCRs found in single-cell data can be found in bulk RNA-Seq data from the same sample. The top CDR3s identified

in single-cell data are mostly identified by all three methods and are identified in the bulk dataset consistently across all methods (**FIG. 32**).

[0095] The rpsTCR pipeline method is comparable to pre-existing methods in datasets with read length of 100bp when sequence assembly is unnecessary. However, in short read datasets where the rpsTCR pipeline implements sequence assembly, sensitivity is greatly improved relative to other methods.

[0096] **FIG. 2** depicts the CDR3 sequences determined from isolated T cells from a tumor of a tumor-implanted mouse or mouse spleen (from the same tumor-bearing mouse) (Example 1). Identical CDR3 sequences detected from 3 or more single T cells were determined to be clonally expanded T cells. The combination treated group (aGITR + aPD-1) had tumors displaying a diverse fraction of clonally expanded T cells by day 11 post treatment (as indicated by the number of different TCR sequences from T cells having 3 or more clones). In this example, the total number of T cell clones that expanded, not the number of a specific TCR sequence, is used to determine clonal expansion after treatment. In some aspects, clonal expansion is determined by the TCR being identical at the amino acid level but the nucleic acid sequence could comprise variations. T cells isolated from spleen had not clonally expanded. Of note are same TCR CDR3 sequences detected in some PD-1 treated mice, yet those were not considered clonally expanded (<3 detected).

[0097] **FIG. 2** illustrates an example TCR sequence library of known TCR sequences and corresponding treatment responses to one or more treatments. As shown in **FIG. 2**, TCR Sequence 1 through TCR Sequence 10 have clonal expansion values of 3 or more in response to treatment by a combination of aGITR and aPD1. A clonal expansion value represents a number of occurrences of a specific TCR sequence among multiple cells, indicating the number of times a cell has been cloned. TCR Sequence 11 through TCR Sequence 17 have clonal expansion values of 3 or more in response to treatment by aPD1 alone. TCR Sequence 18 through TCR Sequence 20 have clonal expansion values of 3 or more in response to treatment by aGITR alone. For a subject whose cell sequence data has been obtained and TCR sequences identified, the identified TCR sequences can be compared to a TCR library as shown in **FIG. 2** and matches determined. In an aspect, the number of matches can be indicative of potential response to a particular treatment. In another aspect, a summation of the clonal expansion values associated for matches on a per treatment

basis can be used to determine which treatment is associated with a higher cumulative clonal expansion value for a subject. By way of example, patient sequence data having TCR Sequence 1, TCR Sequence 2, TCR Sequence 13, TCR Sequence 14, TCR Sequence 15, and TCR Sequence 16 would have a cumulative clonal expansion value of 14 for aGITR and aPD1 in combination and a cumulative clonal expansion value of 16 for aPD1 alone. Thus, the subject is more likely to respond to aPD1 alone.

[0098] **FIG. 3** is a graphical illustration of the results contained in **FIG. 2**. **FIG. 3** is a bar chart indicating that the percentage of expanded clones for aGITR and aPD1 in combination is 31.9% 11 days after treatment, the percentage of expanded clones for aPD1 alone is 26.7% 11 days after treatment (4.8% 8 days after treatment), and the percentage of expanded clones for aGITR alone is 20.3% 11 days after treatment. In an aspect, the results of **FIG. 2** and the chart of **FIG. 3** can be based off of data obtained after administering the one or more treatments to one or more subjects. The method **100** can further comprise administering one or more treatments to the subject. The method **100** can further comprise generating a TCR sequence library of TCR sequences and corresponding clonal expansion associated with the one or more treatments. **FIG. 3** shows the significance of the number of clonally expanded T cells identified in combination treatment of tumor-bearing mice compared to single (mono-) therapy administering either anti-PD1 (aPD1) or anti-GITR (aGITR). Both anti-PD1 (aPD1) and anti-GITR (aGITR) treatments are significantly increased with respect to the percentage of expanded clones compared to tumors from mice treated with isotype control.

[0099] **FIG. 4** illustrates an exemplary output of one embodiment of the method **100** of **FIG. 1**. **FIG. 4** shows T cell lineage after clonal expansion. Within treatment types (e.g., aGITR and aPD1 in combination, aPD1 alone, aGITR alone) clonal expansion occurs across various cell subtypes, such as Naïve, T cell memory (Tcm), T cell effector memory (Tem), T cell effector (Teff), and the like. **FIG. 4** indicates sequences associated with specific treatments along with counts (e.g., clonal expansion values), and includes an indication of cell subtypes associated with the clonal expansion values. **FIG. 4** depicts the profile of T cells in each set of clones. Each T cell clone identified by the TCR CDR3 sequence expressed, was also characterized (e.g., classified by cell surface markers) as a) an effector memory cell (Tem) (Cd62L-Cd44+Gzmb+), b) an effector T cell (Teff) (Cd62L-Cd44-Gzmb+), c)

a naïve T cell (Cd62L+Cd44-Gzmb-), or d) a central memory T cell (Tcm) (Cd62L+Cd44+Gzmb-), or an otherwise unclassified cell (blank oval).

[00100] Disclosed are methods of determining one or more TCR sequences comprising obtaining a first sequence data comprising single cell raw reads from a first cell of a subject, using a bioinformatics tool to map the first sequence data to a second sequence data comprising a plurality of non-T cell receptor transcripts to identify one or more unmapped reads in the first sequence data, and determining one or more TCR sequences from the unmapped reads. In some aspects, obtaining first sequence data comprising single cell raw reads from a first cell of a subject comprises performing random primer RNA sequencing on transcripts obtained from the first cell. The random primers can be 4-40 nucleotides in length. In some instances, the random primers can be 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotides in length. In some aspects, prior to obtaining a first sequence data, the subject is administered an immunotherapy. The immunotherapy can be a monotherapy or a combination therapy. For example, the immunotherapy can be the combination of a costimulatory agonist and a coinhibitory antagonist.

[00101] Disclosed are vectors comprising a CDR3 sequence of a TCR β chain. In some aspects, the vector can be a viral vector or plasmid. Examples of viral vectors can be but are not limited to lentiviral vectors, adenoviral vectors, and adeno-associated viral vectors. In some aspects, the CDR3 sequence is a nucleic acid sequence that encodes CASSRNTEVFF (SEQ ID NO:269), CASSIGNTEVFF (SEQ ID NO:270), CASSQPGKNTTEVFF (SEQ ID NO:271), CASSLGQGNNNSPLYF (SEQ ID NO:272), CASSQGQGAETLYF (SEQ ID NO:273), CASSPPMGGQLYF (SEQ ID NO:274), CASSQEGANTEVFF (SEQ ID NO:275), CASSQVQGTGNTLYF (SEQ ID NO:276), CASSQEGDGYEQYF (SEQ ID NO:277), CTSAEGGGTEVFF (SEQ ID NO:278), CASSPPGGGTEVGG (SEQ ID NO:279), CASSGTDNQDTQYF (SEQ ID NO:280), CASSPGTGGYEQYF (SEQ ID NO:281), CASSLELGFYEQYF (SEQ ID NO:282), CASSLGGAENRERLFF (SEQ ID NO:283), CASSQEGDSYEQYF (SEQ ID NO:284), CASSRNTEVFF (SEQ ID NO:285), CASGDAMGGRDYAEQFF (SEQ ID NO:286), CGAREGQDTQYF (SEQ ID NO:287), CGARTGGEQYF (SEQ ID NO:288), or CTCSAGNQAPLF (SEQ ID NO:289). Thus, in some aspects, disclosed are lentiviral vectors comprising a nucleic acid sequence that encodes the sequence of any one of SEQ ID NOs:269-289.

[00102] In an aspect, TCR sequences with similar binding specificities can be clustered as disclosed in Gupta N.T., et al. Hierarchical clustering can identify B cell clones with high confidence in Ig repertoire sequencing data. *J. Immunol.* 198(6), 2489-2499 (2017). Briefly, CDR3 regions of the TCR sequences can be clustered using single-linkage hierarchical clustering; distance between two CDR3 sequences can be defined as the absolute number of nucleotide differences between the two sequences and a threshold that, in an aspect, can be inferred from the sequence dataset as disclosed in the aforementioned reference.

[00103] Also disclosed are recombinant cells comprising a vector comprising a nucleic acid sequence that encodes the sequence of any one of SEQ ID NOs:269-289.

[00104] Disclosed are recombinant cells comprising a CDR3 sequence of a TCR β chain. In some aspects, the CDR3 sequence is derived from a different cell type, cell line, or different species than the recombinant cell comprising the CDR3 sequence. For example, the CDR3 sequence can be from a primary human T cell and the cell comprising the CDR3 sequence can be a T cell line derived from any other T cell than the cell from which the CDR3 sequence was derived. Another example, the CDR3 sequence can be from a human cell and the cell comprising the CDR3 sequence can be a non-human cell. In some aspects, the recombinant cells comprise a CDR3 sequence comprising the sequence of CASSRNTEVFF (SEQ ID NO:269), CASSIGNTEVFF (SEQ ID NO:270), CASSQPGKNTEVFF (SEQ ID NO:271), CASSLGQGNNSPLYF (SEQ ID NO:272), CASSQGQGAETLYF (SEQ ID NO:273), CASSPPMGGQLYF (SEQ ID NO:274), CASSQEGANTEVFF (SEQ ID NO:275), CASSQVQGTGNTLYF (SEQ ID NO:276), CASSQEGDGYEQYF (SEQ ID NO:277), CTSAEGGGTEVFF (SEQ ID NO:278), CASSPPGGGTEVGG (SEQ ID NO:279), CASSGTDNQDTQYF (SEQ ID NO:280), CASSPGTGGYEQYF (SEQ ID NO:281), CASSLELGFYEQYF (SEQ ID NO:282), CASSLGGAPNERLFF (SEQ ID NO:283), CASSQEGDSYEQYF (SEQ ID NO:284), CASSRNTEVFF (SEQ ID NO:285), CASGDAMGGRDYAEQFF (SEQ ID NO:286), CGAREGQDTQYF (SEQ ID NO:287), CGARTGGEQYF (SEQ ID NO:288), or CTCsAGNQAPLF (SEQ ID NO:289).

[00105] In some aspects, the disclosed methods for identifying TCR sequences from random priming RNA sequencing can be used to identify B cell receptors (BCRs) as

well. The steps of the pipeline are nearly identical except for the following steps: 1) the negative selection step wherein identifying BCRs involves alignment of short reads to a second reference dataset comprising a plurality of species-specific non-B cell receptor RNA transcripts; 2) the assembly step wherein identifying BCRs involves assembly of the one or more unmapped short reads into one or more long reads for further processing can comprise aligning the one or more unmapped short reads to one or more BCR sequences from a reference database of BCR sequences and assembling the one or more unmapped short reads into long reads (candidate BCR sequences) based on the reference database of BCR sequences; 3) the alignment step wherein identifying BCRs involves alignment of candidate BCR sequences to a reference of BCR V and J genes along with identification of the BCR CDR3 region. In an aspect, generating one or more BCR sequences from the one or more long reads can comprise one or more techniques disclosed in Alamyar, E., et al. IMGT tools for the nucleotide analysis of immunoglobulin (IG) and t cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods in Mol. Biol.* 882, 569–604 (2012).

[00106] In an exemplary aspect, the methods and systems can be implemented on a computer **801** as illustrated in **FIG. 8** and described below. Similarly, the methods and systems disclosed can utilize one or more computers to perform one or more functions in one or more locations. **FIG. 8** is a block diagram illustrating an exemplary operating environment for performing the disclosed methods. This exemplary operating environment is only an example of an operating environment and is not intended to suggest any limitation as to the scope of use or functionality of operating environment architecture. Neither should the operating environment be interpreted as having any dependency or requirement relating to any one or combination of components illustrated in the exemplary operating environment.

[00107] The present methods and systems can be operational with numerous other general purpose or special purpose computing system environments or configurations. Examples of well-known computing systems, environments, and/or configurations that can be suitable for use with the systems and methods comprise, but are not limited to, personal computers, server computers, laptop devices, and multiprocessor systems. Additional examples comprise set top boxes, programmable consumer electronics, network PCs, minicomputers, mainframe computers, distributed

computing environments that comprise any of the above systems or devices, and the like.

[00108] The processing of the disclosed methods and systems can be performed by software components. The disclosed systems and methods can be described in the general context of computer-executable instructions, such as program modules, being executed by one or more computers or other devices. Generally, program modules comprise computer code, routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. The disclosed methods can also be practiced in grid-based and distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules can be located in both local and remote computer storage media including memory storage devices.

[00109] Further, one skilled in the art will appreciate that the systems and methods disclosed herein can be implemented via a general-purpose computing device in the form of a computer **801**. The components of the computer **801** can comprise, but are not limited to, one or more processors **803**, a system memory **812**, and a system bus **813** that couples various system components including the one or more processors **803** to the system memory **812**. The system can utilize parallel processing. Parallel processing can be leveraged to perform the disclosed methods. For example, performance of at least a portion of one or more steps of the disclosed methods can be classified as a job. For example, the disclosed methods can be performed for a plurality of samples in parallel. The workload for each job can then be distributed across several processors. A software application can be used to design and run jobs to process data. A job can, for example, extract data from one or more data sources, transform the data, and load it into one or more new locations (e.g., stage the data for processing in another job). In a parallel processing topology, the workload for each job can be distributed across several processors on one or more computers, called compute nodes. In an aspect, the user can modify a configuration file or otherwise interface with software configured to define multiple processing nodes. These nodes work concurrently to complete each job quickly and efficiently. A conductor node computer can orchestrate the work. Parallel processing environments can be categorized as symmetric multiprocessing (SMP) or massively parallel processing

(MPP) systems. In a symmetric multiprocessing (SMP) environment, multiple processors share other hardware resources. For example, multiple processors can share the same memory and disk space, but use a single operating system. The workload for a parallel job is then distributed across the processors in the system. The actual speed at which the job completes might be limited by the shared resources in the system. To scale the system, the number of processors can be increased, memory can be added, or storage can be increased. In a massively parallel processing (MPP) system, many computers can be physically housed in the same chassis. An MPP system can be physically dispersed. In an MPP environment, performance is improved because no resources must be shared among physical computers. To scale the system, computers, along with associated memory and disk resources can be added. In an MPP system, a file system is commonly shared across the network. In this configuration, program files can be shared instead of installed on individual nodes in the system.

[00110] The system bus **813** represents one or more of several possible types of bus structures, including a memory bus or memory controller, a peripheral bus, an accelerated graphics port, or local bus using any of a variety of bus architectures. By way of example, such architectures can comprise an Industry Standard Architecture (ISA) bus, a Micro Channel Architecture (MCA) bus, an Enhanced ISA (EISA) bus, a Video Electronics Standards Association (VESA) local bus, an Accelerated Graphics Port (AGP) bus, and a Peripheral Component Interconnects (PCI), a PCI-Express bus, a Personal Computer Memory Card Industry Association (PCMCIA), Universal Serial Bus (USB) and the like. The bus **813**, and all buses specified in this description can also be implemented over a wired or wireless network connection and each of the subsystems, including the one or more processors **803**, a mass storage device **804**, an operating system **805**, T cell pipeline software **806**, T cell pipeline data **807**, a network adapter **808**, the system memory **812**, an Input/Output Interface **810**, a display adapter **809**, a display device **811**, and a human machine interface **802**, can be contained within one or more remote computing devices **814a,b,c** at physically separate locations, connected through buses of this form, in effect implementing a fully distributed system.

[00111] The computer **801** typically comprises a variety of computer readable media. Exemplary readable media can be any available media that is accessible by the

computer **801** and comprises, for example and not meant to be limiting, both volatile and non-volatile media, removable and non-removable media. The system memory **812** comprises computer readable media in the form of volatile memory, such as random access memory (RAM), and/or non-volatile memory, such as read only memory (ROM). The system memory **812** typically contains data such as the T cell pipeline data **807** and/or program modules such as the operating system **805** and the T cell pipeline software **806** that are immediately accessible to and/or are presently operated on by the one or more processors **803**.

[00112] In another aspect, the computer **801** can also comprise other removable/non-removable, volatile/non-volatile computer storage media. By way of example, **FIG. 8** illustrates the mass storage device **804** which can provide non-volatile storage of computer code, computer readable instructions, data structures, program modules, and other data for the computer **801**. For example and not meant to be limiting, the mass storage device **804** can be a hard disk, a removable magnetic disk, a removable optical disk, magnetic cassettes or other magnetic storage devices, flash memory cards, CD-ROM, digital versatile disks (DVD) or other optical storage, random access memories (RAM), read only memories (ROM), electrically erasable programmable read-only memory (EEPROM), and the like.

[00113] Optionally, any number of program modules can be stored on the mass storage device **804**, including by way of example, the operating system **805** and the T cell pipeline software **806**. Each of the operating system **805** and the T cell pipeline software **806** (or some combination thereof) can comprise elements of the programming and the T cell pipeline software **806**. The T cell pipeline data **807** can also be stored on the mass storage device **804**. The T cell pipeline data **807** can be stored in any of one or more databases known in the art. Examples of such databases comprise, DB2®, Microsoft® Access, Microsoft® SQL Server, Oracle®, mySQL, PostgreSQL, and the like. The databases can be centralized or distributed across multiple systems.

[00114] In another aspect, the user can enter commands and information into the computer **801** via an input device (not shown). Examples of such input devices comprise, but are not limited to, a keyboard, pointing device (*e.g.*, a “mouse”), a microphone, a joystick, a scanner, tactile input devices such as gloves, and other body coverings, and the like. These and other input devices can be connected to the one or

more processors **803** via the human machine interface **802** that is coupled to the system bus **813**, but can be connected by other interface and bus structures, such as a parallel port, game port, an IEEE 1394 Port (also known as a Firewire port), a serial port, or a universal serial bus (USB).

[00115] In yet another aspect, the display device **811** can also be connected to the system bus **813** via an interface, such as the display adapter **809**. It is contemplated that the computer **801** can have more than one display adapter **809** and the computer **801** can have more than one display device **811**. For example, the display device **811** can be a monitor, an LCD (Liquid Crystal Display), or a projector. In addition to the display device **811**, other output peripheral devices can comprise components such as speakers (not shown) and a printer (not shown) which can be connected to the computer **801** via the Input/Output Interface **810**. Any step and/or result of the methods can be output in any form to an output device. Such output can be any form of visual representation, including, but not limited to, textual, graphical, animation, audio, tactile, and the like. The display device **811** and computer **801** can be part of one device, or separate devices.

[00116] The computer **801** can operate in a networked environment using logical connections to one or more remote computing devices **814a,b,c**. By way of example, a remote computing device can be a personal computer, portable computer, smartphone, a server, a router, a network computer, a peer device or other common network node, and so on. Logical connections between the computer **801** and a remote computing device **814a,b,c** can be made via a network **815**, such as a local area network (LAN) and/or a general wide area network (WAN). Such network connections can be through the network adapter **808**. The network adapter **808** can be implemented in both wired and wireless environments. Such networking environments are conventional and commonplace in dwellings, offices, enterprise-wide computer networks, intranets, and the Internet.

[00117] For purposes of illustration, application programs and other executable program components such as the operating system **805** are illustrated herein as discrete blocks, although it is recognized that such programs and components reside at various times in different storage components of the computing device **801**, and are executed by the one or more processors **803** of the computer. An implementation of the T cell pipeline software **806** can be stored on or transmitted across some form of

computer readable media. Any of the disclosed methods can be performed by computer readable instructions embodied on computer readable media. Computer readable media can be any available media that can be accessed by a computer. By way of example and not meant to be limiting, computer readable media can comprise “computer storage media” and “communications media.” “Computer storage media” comprise volatile and non-volatile, removable and non-removable media implemented in any methods or technology for storage of information such as computer readable instructions, data structures, program modules, or other data. Exemplary computer storage media comprises, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by a computer.

[00118] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how the compounds, compositions, articles, devices and/or methods claimed herein are made and evaluated, and are intended to be purely exemplary and are not intended to limit the scope of the methods and systems. Efforts have been made to ensure accuracy with respect to numbers (e.g., amounts, temperature, etc.), but some errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, temperature is in °C or is at ambient temperature, and pressure is at or near atmospheric.

[00119] The methods and systems can employ Artificial Intelligence techniques such as machine learning and iterative learning. Examples of such techniques include, but are not limited to, expert systems, case based reasoning, Bayesian networks, behavior based AI, neural networks, fuzzy systems, evolutionary computation (e.g. genetic algorithms), swarm intelligence (e.g. ant algorithms), and hybrid intelligent systems (e.g. Expert inference rules generated through a neural network or production rules from statistical learning).

[00120] The following examples are provided to describe the embodiments in greater detail. They are intended to illustrate, not to limit, the claimed embodiments.

Example 1

***In vivo* mouse studies**

- [00121] For MC38 tumor studies, 3×10^5 or 5×10^5 MC38 cells were subcutaneously injected on the right flank of C57BL/6 or humanized GITR/GITRL double knock-in mice, respectively (day 0). On day 6 after tumor implantation, mice were grouped based on tumor size and treated by intraperitoneal injection with 5 mg kg⁻¹ anti-GITR (DTA1) and/or anti-PD-1 (RPM1-14) Ab or isotype control IgGs (rat IgG2b, LTF-2 and rat IgG2a, 2A3) at indicated doses (Abs obtained from Bio X Cell). Antibodies were administrated again on day 13. Mice treated with combination of anti-PD-1 (aPD-1) and anti-GITR (aGITR) Ab mice remained tumor-free for over 80 days. These mice were re-challenged with 3×10^5 of MC38 and 2.5×10^5 B16F10.9 cells bilaterally. Naïve mice were used as tumor implantation control.
- [00122] For T cell subset depletion experiments, mice treated with either combination therapy or isotype control IgG were treated with 300 µg depleting mAbs, including anti-CD4, clone GK1.5; anti-CD8 clone 2.43 and rat IgG2b isotype (BioX Cell) and anti-CD25, clone PC61 (eBioscience); rat IgG1 isotype (HPRN, Bio X Cell). Depletion Ab were given at one day prior of tumor challenge (day -1) and twice weekly for total eight doses. The depletion efficiency was confirmed by FACS analysis of peripheral blood samples. Perpendicular tumor diameters were measured blindly 2-3 times per weeks using digital calipers (VWR, Radnor, PA). Volume was calculated using the formula $L \times W \times 0.5$, where L is the longest dimension and W is the perpendicular dimension. Differences in survival were determined for each group by the Kaplan-Meier method and the overall P value was calculated by the log-rank testing using survival analysis by Prism version 6 (GraphPad Software Inc.). An event was defined as death when tumor burden reached the protocol-specified size of 2000 mm³ in maximum tumor volume to minimize morbidity.

Example 2

Single-cell sorting RNA-seq analysis

- [00123] On day 8 and 11 post tumor challenge, single cell suspension of tumor was prepared by mouse tumor dissociation kit (Miltenyi Biotec) and spleens were dissociated with gentleMACS Octo Dissociator. Tumors and spleens from the same treatment group were pooled and viable CD8⁺ T cells were sorted by FACS. FACS sorted T cells were mixed with C1 Cell Suspension Reagent (Fluidigm) before loading onto a 5- to 10-µm C1 Integrated Fluidic Circuit (IFC; Fluidigm). LIVE/DEAD staining solution (Thermo Fisher) was prepared by adding 2.5 µL

ethidium homodimer-1 and 0.625 μ L calcein AM (Life Technologies) to 1.25 mL C1 Cell Wash Buffer (Fluidigm) and 20 μ L was loaded onto the C1 IFC. Each capture site was carefully examined under a Zeiss microscope in bright field, green fluorescent protein (GFP), and Texas Red channels for cell doublets and viability. Cell lysing, reverse transcription, and cDNA amplification were performed on the C1 Single-Cell Auto Prep IFC, as specified by the manufacturer (protocol 100-7168 E1). The SMARTer Ultra Low RNA Kit (Clontech) was used for cDNA synthesis from the single cells. Illumina NGS libraries were constructed using the Nextera XT DNA Sample Prep kit (Illumina), according to the manufacturer's recommendations (protocol 100-7168 E1). A total of 2,222 single cells were sequenced on Illumina NextSeq (Illumina) by multiplexed single-read run with 75 cycles.

[00124] **Table 8: Summary of Captured T cells.**

Treatment	Day post tumor implant	Tissue	# of C1 chips	# of captured cells	# of cells passed QC
Isotype control	8	Tumor	2	98	47
		Spleen	2	102	51
	11	Tumor	3	194	105
		Spleen	2	157	128
	18	Tumor	1	66	
		Spleen	1	69	
aGITR+aPD-1	8	Tumor	2	135	81
		Spleen	2	115	78
	11	Tumor	3	184	141
		Spleen	2	134	113
aGITR alone	8	Tumor	4	255	127
		Spleen	4	240	86
	11	Tumor	2	113	79
		Spleen	1	83	72
aPD1 alone	8	Tumor	2	115	63
		Spleen	1	57	42
	11	Tumor	4	177	120
		Spleen	1	63	46
Total				2,357	1,379

[00125] Raw sequence data (BCL files) from each of these cells were converted to

FASTQ format via Illumina Casava 1.8.2. Reads were decoded based on their barcodes. Read quality was evaluated using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). For TCR analysis, the disclosed methods, including random-priming short-read TCR (rpsTCR) analysis, for reconstructing and extracting TCR sequences, especially TCR-CDR3 sequences from random priming short RNA sequencing reads was used. (See **FIG. 1**). The rpsTCR method provided paired- and single-end short reads and mapped these reads to mouse or human genomes and transcriptomes, but not to TCR gene loci and transcripts, using TopHat43 with default parameters. Thus negative selection allows for mapped reads to be discarded and unmapped reads were recycled for extraction of TCR sequences. Low quality nucleotides in the unmapped reads were trimmed (i.e., reads with length less than 35bp were filtered out using HTQC toolkit). QC passed short reads were assembled into longer reads using iSSAKE default setting. TCRklass was used to identify CDR3 sequences with Scr (conserved residue support score) set from default 1.7 to 2.

Example 3

Analysis of tumor-infiltrating T cells isolated from tumor-bearing mice.

[00126] The disclosed methods (pipeline) were utilized to reconstruct, extract and analyze TCR sequences using single cell sorted RNAseq data, allowing the identification of high-frequency T cell clones potentially associated with tumor reactivity and patient survival. The pipeline was used to profile the transcriptome of 1379 CD8⁺ T cells isolated from tumor-bearing mice. At the early time point (day 8), very few clones of high-frequency T cells (defined as at least 3 T cells sharing identical TCR sequences) were detected in all treatment groups (**FIG. 2**). By day 11, we identified 2 high-frequency T cell clones, representing 5.7% of sequenced single CD8⁺ T cells from isotype control samples; 3 clones/20.3% for anti-GITR samples, 6 clones/26.7% for anti-PD-1 samples and 10 clones/31.9% for combination treated samples. This result indicates that between day 8 and day 11, a strong clonal expansion of intratumoral CD8⁺ T cells was primarily driven by anti-PD-1 treatment. The significance of the number of clonally expanded T cells identified in combination treatment of tumor-bearing mice compared to single (mono-) therapy administering either anti-PD1 (aPD1) or anti-GITR (aGITR) is depicted in **FIG. 3**. Anti-GITR and/or anti-PD-1 had no impact on peripheral/spleen T cell clonality (**FIG. 2**),

consistent with patient data showing that anti-PD-1 (pembrolizumab) treatment did not affect peripheral blood T cell clonality. Although single agent therapy expanded intratumoral CD8⁺ T cell clones and modulated critical gene pathways, it was not sufficient for complete tumor rejection. The data suggests that a profound reprogramming of dysfunctional tumor infiltrating T cells by combination therapy was required for complete tumor rejection.

Example 4

T cell activation assay for tumor derived TCRs.

[00127] T cell activation assays measured by Luciferase expression in JRT3 cell lines expressing isolated TCRs comprising identified CDR3 sequences (**FIGs. 6A-6D**). TCR comprising CDR3 sequence (TCR Sequence Seq.17) isolated from an anti-GITR treated tumor T cell that was highly expanded (i.e., 10 clones identified) was expressed in a JRT3 cell line. The JRT3 becomes activated in the presence of tumor expressing the original antigen, MC38 (**FIG. 6A**), indicating that the transfected cell line recognizes the antigen. TCR Sequence Seq.17-expressing cell lines also become activated when anti-CD3/anti-CD28 stimulated (**FIG. 6B**). Analogous studies using TCR Sequence Seq.12-expressing cell lines (TCR CDR3 sequence was isolated from a aPD-1 stimulated T cell having a Clone size of 8). The TCR Sequence Seq.12-expressing cell line was activated in the presence of tumor expressing the original antigen (**FIG. 6C**) and was activated in the presence of anti-CD3/anti-CD28 (**FIG. 6D**).

Example 5

Gene expression analysis

[00128] Gene expression analysis tools were also utilized to profile the transcriptome (mapped portion of sequences) of the 1379 CD8⁺ T cells isolated from tumor-bearing mice. To identify unique gene signatures in clonal expanded CD8⁺ T cells from combination treatment samples, comparisons across treatment groups were performed. T cell lineage after clonal expansion was identified by the TCR CDR3 sequence expressed, correlated to the expression pattern of cell surface markers. See **FIG. 4**. As such, clonal T cells were classified as effector memory cells (Tem) (Cd62L-Cd44-Gzmb+), effector T cells (Teff) (Cd62L-Cd44-Gzmb+) (or unclassified cells, i.e. non-clonally expanded if less than 3 T cells having the expressed TCR, such as naïve T cells (Cd62L+Cd44-Gzmb-) or central memory T

cells (Tcm) (Cd62L+Cd44+Gzmb-)).

[00129] Gene expression analysis also yielded differentially expressed gene profiles across the clonally expanded and nonclonally expanded cells in each of the treatment groups. *CD226* was identified as one of the two genes shared across different comparison pairs (**FIG. 5**). **FIG. 5** is a Venn diagram illustrating genes preferably expressed in expanded intratumoral CD8⁺ T cells of aGITR+aPD1 treatment group on day 11 of treatment. The Venn diagram summarizes the gene signature analysis of clonal expanded/non-expanded CD8⁺ T cells among treatment groups. *CD226* and *PDE4D* were identified as two genes shared across different comparison pairs.

[00130] **FIG. 5** indicates that two genes are preferably expressed when comparing genes expressed under treatment with aGITR and aPD1 in combination to all other treatments. The two genes are *CD226* and *Pde4d*. The table underneath the Venn diagram provides the occurrences of each gene when comparing genes expressed under treatment with aGITR and aPD1 in combination to all other treatments. Thus the disclosed method can be used to identify specific genes that play a role in immune response to tumor cells.

[00131] *CD226* is a costimulatory molecule that plays an important role in anti-tumor response. Expression analysis in different subsets of intratumoral CD8⁺ T cells (total, clonal expanded, or non-expanded) across treatment groups revealed that *CD226* mRNA levels were significantly increased by combination treatment on clonal expanded T cells (**FIG. 7A**), while this difference was diluted in total and non-expanded CD8⁺ T cells. This observation stresses the importance of performing genome profiling on putative tumor-reactive clones (high-frequency T cell clones) to unmask critical gene changes, and also allows for identification of biomarkers that are informative about efficacious treatments that affect T cell activity.

[00132] **FIG. 7A** illustrates cumulative distribution function (CDF) plots showing expression of a key regulated gene, *CD226*, in total, clonal expanded or non-expanded CD8 T cells. The clonal expanded CD8 cells show the highest expression of *CD226*. The disclosed method is useful to classify subjects that show better correlation with or expression of particular genes. These data also show the utility of the method to identify prognostic genes that inform whether expression is correlated with T cell expansion during a particular treatment, and therefore indicates likelihood of efficacy. *CD226* significantly correlates with improved survival when analyzed from the T

cells of patients with melanoma, lung squamous cell carcinoma and sarcoma (**FIG. 7B**). **FIG. 7B** illustrates TCGA data analysis of CD226 expression level and overall survival in patients with melanoma, lung squamous cell carcinoma and sarcoma. (*, $P < 0.05$, **, $P < 0.01$; ***, $P < 0.001$, ****, $P < 0.0001$ between selected relevant comparison). As shown, CD226 correlates with improved survival.

Example 6A

CD226/TIGIT axis mediates durable anti-tumor responses upon PD-1 and GITR combination immunotherapy

Introduction

[00133] Single or combination therapy targeting immune checkpoints PD-1 and CTLA-4 shows significant clinical benefit in certain cancer patient populations. However, the majority of patients either are resistant or only respond transiently, raising fundamental questions about the selection of optimal immune-modulatory targets to address patient-specific tumor sensitivity. Combination treatment targeting specific coinhibitory and costimulatory pathways to induce a stronger T cell activation, can lead to more durable anti-tumor responses. Here, PD1 and GITR combination therapy, a pre-clinically validated modality currently in early phase clinical testing, was used to characterize the molecular pathways driving long-term responses. Single cell RNA-seq libraries prepared from over 2,000 tumor infiltrating CD8⁺ T cells were sequenced and found that the combination of GITR and PD-1 antibodies synergistically enhanced CD8⁺ T cell effector function by restoring the balance of key homeostatic regulators CD226 and TIGIT, resulting in significant survival benefit. Indeed, anti-PD-1 treatment enhanced CD226 cell surface expression. However, PD-1 monotherapy was insufficient to overcome the inhibitory signaling mediated by TIGIT. Anti-GITR antibody decreased TIGIT expression on T cells. Thus, combination therapy synergistically regulated the strength of CD8⁺ T cell response, and elicited potent adaptive immunity. Indeed, costimulation via CD226 is essential for anti-tumor immunity as genetic inactivation or pharmacological inhibition of CD226 reversed the tumor regression mediated by combination treatment, while inhibition of other TNF-receptor or B7 superfamily members had no effect. Importantly, RNA-seq analysis on tumor biopsies from 43 advanced cancer patients pre and post-anti-PD-1 therapy revealed that CD226 expression was significantly increased after anti-PD-1 treatment. Further high levels of CD226 were

correlated with better prognosis in patients with different types of cancer. Such biomarkers in addition to PD-1/PDL-1 could improve patient selection. Systematic approaches unmasking the molecular pathways driving durable anti-tumor responses by rebalancing homeostatic regulators can be important to optimize combination immunotherapy.

[00134] Following the clinical success of PD-1 and CTLA4 antibody treatments, the therapeutic arsenal of agents in immunotherapy is expanding rapidly. A key goal is to improve the limited response rate and/or the durability of the anti-tumor response achieved with monotherapy approaches in cancer patients. Combination treatments targeting specific coinhibitory (PD-1) and costimulatory (GITR, glucocorticoid-induced TNFR-related protein, *TNFRSF18*) pathways inducing a stronger T cell activation are currently being evaluated in early phase clinical trials for patients with metastatic melanoma and other solid tumors. Indeed, the clinical relevance of T cells in the control of a diverse set of human cancers is now beyond doubt. GITR is constitutively expressed at a high level on T_{reg} cells and can be induced on other lymphocytes upon activation. DTA-1, an agonistic anti-mouse GITR Ab reduces intratumoral T_{reg} cells and mediates FcγR-dependent tumor rejection. Additionally, engaging GITR receptor with an agonistic Ab delivers costimulatory signals directly to effector T cells. While anti-GITR and anti-PD-1 Ab monotherapy has limited efficacy in large or poorly immunogenic tumors, combination therapy promotes long-term survival in ovarian and breast tumor models. However, the molecular mechanism underlying the synergism remains unknown. Here, PD1 and GITR combination therapy was used and over 2000 tumor infiltrating CD8⁺ T cells in a murine MC38 colon adenocarcinoma model were genetically profiled. The systematic approach unmasked the molecular pathways driving durable anti-tumor responses, providing a basis by which to optimize existing combination immunotherapies, and identify new potential biomarkers to improve patient stratification and tumor sensitivity.

Methods

[00135] *Cell lines and tissue culture.* MC38 mouse colon carcinoma cells and RENCA mouse renal adenocarcinoma cells were obtained from American Type Culture Collection (ATCC) and were cultured at 37°C, 5% CO₂ in DMEM media supplied with 10% FBS, 100U mL⁻¹ penicillin and 100 µg mL⁻¹ streptomycin, 2 mM L-

glutamine, 100 μ M NEAA (ThermoFisher Scientific). J.RT3-T3.5 mutant Jurkat cell line lacking endogenous TCR expression was obtained from ATCC and maintained in RPMI-1640 media with 10% FBS. Tumor cell lines were tested negative for Mycoplasma and common rodent pathogens by IMPACT test. MC38- OVA- β_2 m- K^b were generated by transducing MC38 tumor cells with lentiviral vector (LV) encoding a single trimer consisting of SIINFEKL peptide-spacer- β_2 microglobulin-spacer MHC class I (K^b) heavy chain. Surface expression of single trimer was confirmed with 25D-1.16 Ab (eBioscience, **FIG. 21A**). MC38- OVA- β_2 m- K^b were maintained with selection media containing 1.25 μ g ml⁻¹ puromycin (ThermoFisher Scientific). MC38 tumor cells over-expressing CD155 were generated by transduction with LV encoding mouse full length CD155 and FACS sorted on the top 5% expressing cells. Expression level of CD155 was confirmed by FACS analysis (**FIG. 24C**).

[00136] *Mice.* Six to eight week old female C57BL/6 mice were obtained from The Jackson Laboratory. CD226^{-/-} and TIGIT^{-/-} mice in C57BL/6 background were generated at Regeneron using the VelociGene® method. Briefly, EGFP (for CD226) or LacZ cDNA (for TIGIT) was inserted in-frame to the start codon, followed by a selection cassette which disrupts transcription of the gene body and results in a CD226 or TIGIT null allele. Heterozygous targeted mice were interbred to produce homozygous knockout mice for study. All animals were maintained under pathogen-free conditions and experiments were performed according to protocols approved by the Institute of Animal Care and Use Committee (IACUC) of Regeneron Pharmaceuticals, Inc.

[00137] *In vivo mouse studies.* For MC38 tumor studies, 3x10⁵ MC38 cells were subcutaneously injected on the right flank of age-matched C57BL/6 mice (day 0). On day 6 after tumor implantation, mice (randomly distributed in different groups) were grouped based on tumor size and treated by intraperitoneal injection with 5 mg kg⁻¹ anti-GITR (DTA1) and/or anti-PD-1 (RPM1-14) Ab or isotype control IgGs (rat IgG2b, LTF-2 and rat IgG2a, 2A3) at indicated doses (antibodies were obtained from Bio X Cell). Antibodies were administered again on day 13. For antibody depletion experiments, mice treated with either combination therapy or isotype control IgG were treated with 300 μ g depleting or isotype control mAbs, including anti-CD4 (clone GK1.5); anti-CD8 (clone 2.43) and rat IgG2b isotype (clone LTF-2), rat IgG1 isotype (clone HPRN, Bio X Cell) and anti-CD25 (clone PC61, eBioscience).

Depletion Ab were given at one day prior of tumor challenge (day -1) and twice weekly for total eight doses. The depletion efficiency was confirmed by FACS analysis of peripheral blood samples (**FIG. 16C**). Blocking antibodies used in this study include anti-CD226 Ab (clone 10E5, rat IgG2b, eBioscience, 25 mg/kg), CD28 blocking (CTLA4-Fc, Orencia, BMS, 10 mg/kg), anti-OX40L (clone RM134L, rat IgG2b, Bio X Cell, 10 mg/kg) and anti-4-1BBL (clone TKS-1, rat IgG2a, Bio X Cell, 10 mg/kg). Blocking Ab were given twice weekly by i.p. injection starting 1-2 days prior to immunotherapy, for two weeks. Perpendicular tumor diameters were measured blindly 2-3 times per weeks using digital calipers (VWR, Radnor, PA). Volume was calculated using the formula $L \times W \times 0.5$, where L is the longest dimension and W is the perpendicular dimension. Differences in survival were determined for each group by the Kaplan-Meier method and the overall p value was calculated by the log-rank testing using survival analysis by Prism version 6 (GraphPad Software Inc.). An event was defined as death when tumor burden reached the protocol-specified size of 2000 mm³ in maximum tumor volume to minimize morbidity.

[00138] *Flow cytometry.* For flow cytometry analysis of in vivo experiments, blood, spleen, thymus, lymph node and tumor were harvested on indicated days post treatment. Single cell suspensions were prepared and red blood cells were lysed using ACK Lysis buffer (ThermoFisher Scientific). Live/dead cell discrimination was performed using Live/dead fixable blue dead cell staining kit (ThermoFisher Scientific). Cells were first stained with Abs for surface markers for 20-30 min at 4°C. Intracellular staining was done using a fixation/permeabilization kit (eBioscience). To quantify OVA-specific CD8 T-cells, single cell suspension was first stained with H-2Kb/SIINFEKL-Pentamer (ProImmune) for 10 min at room temperature before surface markers staining. For intracellular cytokine staining (ICS), cells were stimulated with or without SIINFEKL peptide for 36 hours and with Protein Transport Inhibitor (BD Bioscience) for the last 4 hours. After stimulation, cells were stained as described above for surface and intracellular proteins. To quantify cell numbers in tissue, a fixed number of CountBright Absolute Counting Beads (ThermoFisher Scientific) were added to each sample prior to acquiring. Samples were acquired on Fortessa X20 or LSR II (BD Bioscience) and analyzed using FlowJo software (TreeStar). See Supplementary Methods for a list of antibodies used.

[00139] *Single-cell sorting RNA-seq analysis.* On day 8 and 11 post tumor challenge, single cell suspensions of tumor were prepared using a mouse tumor dissociation kit (Miltenyi Biotec) and spleens were dissociated with gentle MACS Octo Dissociator. Tumors and spleens from the same treatment group were pooled and viable CD8⁺ T cells were sorted by FACS. FACS sorted T cells were mixed with C1 Cell Suspension Reagent (Fluidigm) before loading onto a 5- to 10- μ m C1 Integrated Fluidic Circuit (IFC; Fluidigm). LIVE/DEAD staining solution was prepared by adding 2.5 μ L ethidium homodimer-1 and 0.625 μ L calcein AM (Life Technologies) to 1.25 mL C1 Cell Wash Buffer (Fluidigm) and 20 μ L was loaded onto the C1 IFC. Each capture site was carefully examined under a Zeiss microscope in bright field, GFP, and Texas Red channels for cell doublets and viability. Cell lysing, reverse transcription, and cDNA amplification were performed on the C1 Single-Cell Auto Prep IFC, as specified by the manufacturer (protocol 100-7168 E1). The SMARTer Ultra Low RNA Kit (Clontech) was used for cDNA synthesis from the single cells. Illumina NGS libraries were constructed using the Nextera XT DNA Sample Prep kit (Illumina), according to the manufacturer's recommendations (protocol 100-7168 E1). A total of 2,222 single cells were sequenced on Illumina NextSeq (Illumina) by multiplexed single-read run with 75 cycles. Raw sequence data (BCL files) were converted to FASTQ format via Illumina Casava 1.8.2. Reads were decoded based on their barcodes. Read quality was evaluated using FastQC (bioinformatics. babraham.ac.uk/projects/fastqc/).

[00140] *Large Unilamellar Vesicles (LUVs).* Phospholipids (79.7% POPC + 10% POPS + 10% DGS-NTA-Ni + 0.3% Rhodamine-PE) were dried under a stream of Argon, desiccated for at least 1 hour and suspended in 1x Reaction buffer (50 mM HEPES-NaOH, pH 7.5, 150 mM NaCl, 10 mM MgCl₂, 1 mM TCEP). LUVs were prepared by extrusion 20 times through a pair of polycarbonate filters with a pore size of 200 nm, as described previously.

[00141] *LUV Reconstitution and Phosphotyrosine Western Blot.* Proteins of interest were pre-mixed at desired ratios in 1x Reaction Buffer containing 0.5 mg/ml BSA, and then mixed with LUVs (1 mM total lipids). The proteins-LUVs mixture incubated at room temperature for 1 hour, during which the His-tagged proteins bound to the liposomes whereas other proteins remained in the extravesicular solution. 2 mM ATP was then injected and rapidly mixed, to trigger phosphorylation, dephosphorylation

and protein interactions at the membrane surface. The reactions were allowed to proceed at room temperature for 30-60 min, and terminated with SDS sample buffer. The samples were heated at 95 °C for 5 min, and subjected to SDS-PAGE. Proteins were transferred to nitrocellulose membranes using iBlot™ Dry Blotting system (ThermoFisher Scientific). The membranes were blocked with 5%BSA in Tris-buffered saline (pH 7.4) with 0.1% Tween-20, incubated with desired phosphotyrosine specific antibodies, and detected with HRP based enhanced chemiluminescence. The following primary antibodies used: anti-pY142-CD3ζ (BD Biosciences #558402), anti-pY20 (Santa Cruz Biotechnology #sc-1624, for detection of tyrosine phosphorylated CD28 in reconstitution assays), anti-pY418-Src (BD Biosciences #560095, for detection of pY394-Lck), anti-pY505-Lck (Cell Signaling #2751), anti-pY315-ZAP70 (Abcam #ab60970), anti-pY493-ZAP70 (Cell Signaling #2704).

[00142] *Clinical biopsies handling, RNA extraction and RNA-seq.* Biopsies were homogenized in at least 600uLs RLTPlus, with mercaptoethanol added (Sigma Aldrich), on the Omni Shredder (Omni-Inc) for 1 minute at 22,000 RPM. RNA and DNA were extracted using the Qiagen Allprep DNA/RNA Mini Kit (Qiagen) according to the manufactures instructions in the “AllPrep DNA/RNA Mini Handbook” (November 2005) using the protocol on page 26 “Protocol: Simultaneous Purification of Genomic DNA and Total RNA from Animal Tissues.” The optional DNase digestion outlined in Appendix E was used during RNA extraction. An additional 500uL 70% ethanol wash with a 2-minute spin was run after the Buffer AW2 wash, but before the last drying spin, to remove excess salts from the DNA extraction. RNA was quantified on the Nanodrop (ThermoFisher Scientific), and quality was assessed on the Fragment Analyzer (Advanced Analytical) with the ‘Standard Sensitivity RNA Analysis Kit’ (Advanced Analytical) according to the manufacture’s protocol. DNA was quantified with the Qubit dsDNA BR Assay Kit (ThermoFisher Scientific) on the Infinite M200 Pro (Tecan) according to the custom protocol ‘Using the Tecan Microplate Reader for DNA Quantification (BR dsDNA Assay). Completed samples were stored at -80°C in barcoded screw cap tubes. For RNA-seq, strand-specific RNA-seq libraries were prepared from 100 ng total RNA using KAPA stranded mRNA-Seq Kit (KAPA Biosystems) and the libraries with size between 400 to 600 bp were selected using Pippin system (Sage Science). Pair end

2x100 bp sequencing was done using Illumina 2500. RNA-seq reads was QCed and aligned to the reference genome and gene expression was quantitated using Array Studio (Omicsoft).

[00143] *Statistical Analysis.* Sample sizes were chosen empirically to ensure adequate statistical power and were in the line with field standards for the techniques employed in the study. Statistical significance was determined with ANOVA or un-paired two-tailed Student's t-test assuming unequal variance at $P < 0.05$ level of significance (or indicated in figure legends).

Results

[00144] To examine the effect of combination immunotherapy poorly immunogenic tumor models (MC38 and RENCA) were used. Although variable reduction of tumor volume and modestly prolonged survival have been reported, monotherapy with anti-PD-1 or anti-GITR Ab is not effective at inducing complete and durable tumor regression in established tumors. Here, antibodies were administrated 6 and 13 days post-tumor challenge when tumors were palpable. Consistent with published data, anti-GITR or anti-PD-1 treatment alone showed no or little effect. Combination therapy synergistically eradicated tumors in the majority (12 tumor free out of 17) of the mice (**FIG. 16A**) and promoted long-term survival (~70% of mice were tumor-free for >80 days) (**FIG. 16B**), in a CD8⁺ T cell-dependent manner (**FIG. 16C** and **16D**), and increased the ratio of intratumoral CD8⁺/T_{reg} and CD4⁺ T effector (T_{eff})/T_{reg} cells (**FIG. 16E, 16F**) in agreement with previous data. After 50 days, only 0-2 of 17 mice were tumor free and 0-10% were alive in monotherapy treated groups (**FIG. 16A, 16B**). Further, the dysfunctional state of the intratumoral CD8⁺ T cells was significantly reversed only upon combination treatment, as indicated by restored *ex vivo* proliferation potential (expression of Ki67, **FIG. 16G**) and effector function (expression of granzyme A and granzyme B, **FIG. 16H** and **16I**). The synergistic anti-tumor effect of combination treatment associated with better survival rate was also confirmed in a second mouse RENCA tumor model (**FIG. 16J**).

[00145] To identify unique gene signatures in clonally expanded CD8⁺ T cells (tumors harvested at day 11) from combination treatment samples, comprehensive comparisons were performed across different treatment groups. First, an RNA signature change in 30 genes after combination treatment was observed, which it was even more significant within the expanded CD8 T cell population (**FIG. 10A** and

FIG. 20). Next, a four-way comparison was performed across all groups to identify genes specifically regulated upon combination versus monotherapy treatment (**FIG. 10B**). *CD226* was identified as one of the two genes shared across different comparison pairs (**FIG. 10B**). *CD226* is a costimulatory molecule that plays an important role in anti-tumor response. Expression analysis of different subsets of intratumoral CD8⁺ T cells (total, clonally expanded, or non-expanded) across treatment groups (**FIG. 10C**) revealed that *CD226* mRNA levels were significantly increased by combination treatment on clonally expanded T cells (fold change = 10.7), while this difference was diluted in bulk (fold change = 3.5) and non-expanded CD8⁺ T cells (not significant) (**FIG. 10D**). Further, *CD226* mRNA levels were significantly increased by combination treatment on clonally expanded CD8 T cells in comparison to anti-PD-1 (fold change = 6.5) and anti-GITR (fold change = 9.2) (**FIG. 10D**). This observation stresses the importance of performing genome profiling on putative tumor-reactive clones (high-frequency T cell clones) to unmask critical gene changes.

[00146] To evaluate the expression levels of *CD226* on intratumoral CD8 T cells after combination treatment MC38 specific TCR clones were tracked *in vivo* using recently published mutated MC38 tumor epitopes. This approach was not successful. The inability of these T cell clones to recognize previously characterized MC38 tumor neo-epitopes could reflect the different mutation status of tumor cell lines between laboratories, likely due to genome instability of the tumor cells. To functionally validate the findings, an MC38 tumor cell line expressing H-2Kb single-chain trimer of MHC class I with SIINFEKL peptide and β_2m (OVA- β_2m -K^b) was generated (**FIG. 21A**), and used SIINFEKL as a surrogate tumor epitope. Consistent with the TCR clonality analysis, anti-PD-1 Ab or combination treatment with anti-GITR induced significant clonal expansion of OVA-specific CD8⁺ T cells (using K^b/OVA pentamer staining) in tumor infiltrating T cells, but only the combination treatment significantly increased the intratumoral density of the Ag-specific CD8⁺ T cell clones (**FIG. 21B**). Further, only combination treatment significantly expanded OVA-specific CD8⁺ T cells in spleen (**FIG. 21B**) compared to control groups, thus extending previous findings. These clonally expanded OVA-specific T cells were functional and produced higher level of IFN γ upon restimulation with OVA peptide than controls (**FIG. 21C**). Importantly, using the MC38-OVA- β_2m -K^b model, it was found that

baseline levels of CD226 were highest on spleen OVA-specific CD8⁺ T cells after anti-PD1 treatment (**FIG. 11A**) and were further elevated by combination treatment with anti-GITR and anti-PD-1 Ab. The same treatment had no significant effect on the CD226 levels of non-specific CD8⁺ T cells. Overall anti-PD-1 treatment played a dominant role in driving the increase of CD226 (**FIG. 11A**) providing key information on the mode of action of anti-PD-1 in anti-tumor immunity.

[00147] Next, an association between PD1 and CD226 molecules was investigated. Recent data demonstrated a highly specific recruitment of Shp2 by PD1 using Fluorescence Energy Transfer (FRET)-based assay in a cell-free reconstitution system in which cytoplasmic domain of PD1 was bound to the surface of large unilamellar vesicles (LUVs) that mimic the plasma membrane of T cells. To examine if CD226 is a target for desphosphorylation by the PD1-Shp2 complex different components (CD3, CD226, and legend/method) involved in T cell signaling were reconstituted on the liposomes (**FIG. 11B**). The sensitivity of each component in response to PD-1 titration on the LUVs was measured by phosphotyrosine (pY) western blots. Previous published data showing that TCR/CD3 ζ was not a sensitive target to desphosphorylation by PD-1-Shp2 was confirmed (**FIG. 11C**). Importantly, it was found that CD226 was very efficiently dephosphorylated by PD1-Shp2 in a dose dependent manner (**FIG. 11C**). The data demonstrate an association between PD-1 and CD226.

[00148] It has been recently shown that the strength of CD8⁺ T cell response is impacted by the overall balance between CD226 and co-inhibitory receptor TIGIT. Interestingly, using single cell RNA-seq it was found that anti-GITR Ab treatment increased TIGIT transcripts in high-frequency T cell clones (**FIG. 22A**), while FACS analysis showed a significant decrease in expression of TIGIT on OVA-specific CD8⁺ T cells (**FIG. 22B**). This result is consistent with the previous finding that TIGIT expression is tightly regulated at the post-transcriptional level. Both cis- and trans-inhibitory mechanisms have been proposed for the TIGIT/CD226 signaling pathway, therefore the net outcome can result from the balance between the expression level of CD226 on CD8⁺ T cells and TIGIT on both CD8⁺ T cells and stand by lymphocytes. Indeed, combination treatment significantly decreased the percentage of TIGIT⁺ cells and the expression level on a per cell basis on bulk tumor infiltrating CD8⁺, CD4⁺ T_{eff} and T_{reg} cells, the effect of which was mainly driven by anti-GITR Ab treatment

(**FIG. 11D** and **FIG. 22C**). Surprisingly, this effect was also found in spleen T cell subsets (**FIG. 11D** and **FIG. 22C**). Combination and/or monotherapy treatment had no effect on bulk CD226⁺ tumor infiltrating or splenic CD4⁺ and T_{reg} cells (**FIG. 22D**). Overall single-cell sorted RNA-seq and FACS phenotyping data showed that anti-PD-1 favored the expression of CD226, while anti-GITR treatment down-regulated surface expression of TIGIT, synergistically restoring the homeostatic T cell function.

[00149] Using a CD226 blocking mAb, it was shown that costimulatory signaling through CD226 is required for the anti-tumor immunity mediated by combination treatment (**FIG. 12A**). However, as CD226 Ab could have a potential depleting effect on subset of CD8 T cells, CD226 was genetically inactivated in C57BL/6 background mice to repeat this study (**FIG. 23A, B**). CD226^{-/-} mice showed no defect on T cell (CD4⁺, CD8⁺, T_{regs}) homeostasis (**FIG. 23C-E**) and responded similarly to wild-type mice to TCR activation (**FIG. 23F**). Importantly, it was found that combination treatment no longer conferred anti-tumor effect or survival benefit in CD226^{-/-} mice, indicating that CD226 was essential for observed anti-tumor effects of the combination (**FIG. 12B**). In addition, the specificity of the CD226 pathway mediating the effect was validated, as inhibition of other members of the TNF receptor superfamily (OX40L or 4-1BBL) or blockade of the B7 costimulatory molecule (CD28) using CTLA4-Ig preserved the anti-tumor effect mediated by the combination therapy (**FIG 12C-E**).

[00150] Further, the CD226 signaling pathway was required for enhanced tumor surveillance in TIGIT^{-/-} mice (**FIG. 24A, B**). Interestingly, mice bearing MC38 tumor cells overexpressing the major ligand for CD226, CD155/PVR (**FIG. 24C**) showed significant delay of tumor growth upon anti-PD-1 or anti-GITR or combination therapy in comparison to MC38-empty vector (MC38-EV) tumor cells or mice treated with isotype control (**FIG. 24D**). Immune profiling analysis of mice transplanted with MC38-CD155 confirmed sustained higher CD155 expression level on MC38-CD155 cells over M38-EV (empty vector) post-implantation (**FIG. 24E**). CD155 over-expression on MC38 tumor cells was associated with decreased detectable CD226 expression on CD4⁺, CD8⁺ T and T_{regs} cells, while it boosted T cell activation as indicated by enhanced IFN γ and 4-1BB expression on intra-tumoral T cells (**FIG. 24F, 24G**). As expected, no effect was observed in the periphery.

[00151] Next, the relationship between PD-1 inhibition and CD226 expression was investigated in a clinical setting. RNA-seq analysis was performed on tumor biopsies collected from 43 advanced cancer patients pre- and post-PD-1 targeted treatment (**FIG. 12F**). Importantly, CD226 expression was significantly increased after two doses of anti-hPD-1 treatment in cancer patients (**FIG. 12G**). Further, clinical data from The Cancer Genome Atlas (TCGA) was interrogated to examine if CD226 expression level correlates with the overall T cell activation strength and can be predictive of a better prognosis in cancer patients. Indeed, patients with high baseline CD226 expression have significantly higher survival probabilities in five (skin cutaneous melanoma, lung adenocarcinoma, head and neck squamous carcinoma, uterine corpus endometrial carcinoma and sarcoma) out of twenty different types of cancer evaluated here (**FIG. 12H** and **FIG. 25**). Overall, these results support an immunotherapy strategy that boosts CD226 signaling while blocking TIGIT simultaneously for maximum T cell activation.

[00152] Here, the use of technology platforms to unveil molecular mechanisms driving the potent synergism of a costimulatory agonist and a coinhibitory antagonist, elucidated the parameters required for durable anti-tumor responses and shed light on key functional T cell regulatory pathways that could shape the next generation of tumor specific combination therapies.

Example 6B

TCR repertoire analysis bioinformatics pipeline rpsTCR and its validation

[00153] *TCR sequence extraction and assembly.* Given the V and J allele information, and the CDR3 amino acid sequence, the amino acid sequences of the V and J alleles was extracted from the IMGT database (imgt.org). Next, the CDR3 sequence were aligned with the C-terminal of the V sequence and the N-terminal of the J sequence, to create a contiguous VDJ amino acid sequence. For each V allele, the leader sequence(L) was then identified from IMGT if it is available and appended it to the C-terminal of the VDJ sequence. If the leader sequence was not available, then the most frequent leader sequence was used. The LVDJ amino acid sequence was then back-translated to a codon-optimized nucleotide sequence using the EMBOSS Backtranseq tool (ebi.ac.uk/Tools/st/emboss_backtranseq). Finally, the nucleotide sequences of the constant (C) regions of the TCRA/TCRB (derived from IMGT) were appended to the N-terminal of the LVDJ nucleotide sequence, and thus obtained the full LVDJC

sequences for cloning.

[00154] A bioinformatics pipeline was developed and validated to extract, reconstruct and analyze TCR sequences using random priming RNAseq data generated from sorted single cells allowing the identification of T cell clones potentially associated with tumor reactivity and patient survival. Unlike conventional TCR-seq methods using targeted TCR amplicon sequencing with long reads (2x300bp), a very small portion of random priming RNA-seq reads are TCR sequences and the read length is short (usually ≤ 100 bp), which usually only covers part of the V(D)J regions of the TCRs. To address these issues, a negative TCR sequence selection step was integrated and a short read assembly step in the pipeline. In brief, the pipeline takes paired or single-end short reads and maps these reads to human or mouse genomes and transcriptomes, but not TCR gene loci and transcripts (**FIG. 17** and Methods). The results indicated that the method is a highly sensitive and accurate CDR3 assembler for random priming RNA-seq data (**FIG. 18**, **FIG. 13** and Example 6A Methods, above). Finally, the pipeline was applied to a library of 1,379 CD8⁺ single cell RNA-seq data. The detection rates of TCRB-CDR3 (86%), TCRA-CDR3 (78.2%) and paired TCRB and TCRA (73.1%) were comparable to the reported detection rates using targeted TCR sequencing from single T cells (**FIG. 14**).

[00155] Antibodies were administered 6 days post-tumor challenge when tumors were palpable (**FIG. 9A**). The above bioinformatics tool was used to profile the transcriptome of 1379 CD8⁺ T cells single cell sorted from tumor-bearing mice at day 8 and day 11 post-tumor challenge. At the early time point (day 8), very few clones of high-frequency T cells (defined as at least 3 T cells sharing identical TCR sequences) were detected in all treatment groups (**FIG. 9B**). By day 11, two high-frequency T cell clones were identified, representing 5.7% of sequenced single CD8⁺ T cells from isotype control samples; 3 clones/20.3% for anti-GITR samples, 6 clones/26.7% for anti-PD-1 samples and 10 clones/31.9% for combination treated samples. Indeed, between day 8 and day 11, a significant clonal expansion of intratumoral CD8⁺ T cells was induced by anti-PD-1 monotherapy. This result is in agreement with published data detecting an increase in TCR clonality after anti-PD-1 therapy (pembrolizumab). The results extend these findings by showing that anti-PD-1 Ab seems to be the main driver of clonal expansion, and that dual targeting of PD-1 and GITR further enhances intra-tumoral CD8 T cell TCR clonality (**FIG. 9B** and 9C). Anti-GITR and/or anti-

PD-1 had no impact on peripheral/spleen T cell clonality (**FIG. 9C**), consistent with patient data using pembrolizumab treatment.

[00156] To validate the tumor antigen specificity of the TCRs enriched within the MC38 tumors upon combination treatment, bioinformatics analysis was performed to extract and assemble the full length paired TCR alpha/beta sequences (Example 6A Methods). Full length TCR pairs derived from expanded CD8⁺ T cells were cloned into lentiviral constructs and transduced into a Jurkat T cell line lacking endogenous TCR expression. AP-1 driven luciferase reporter was used as a read-out of TCR specificity of these engineered T cell lines (**FIG. 9D**). This assay was validated using a well-characterized LCMV TCR (P14) (**FIG. 19**). The TCRs from high frequency clones (5 cells sharing the same TCR) can recognize epitopes that are presented specifically by wild type MC38 tumor cells but not by irrelevant syngeneic C57BL/6 tumor cell lines (B16F10.9 or TRAMP-C2) (**FIG. 9E** showing a representative TCR clone), thus validating their specificity.

[00157] Further, it was determined that anti-GITR and anti-PD-1 regulate distinct molecular pathways in these clonally expanded CD8⁺ T cells, (**FIG. 15**). Combination treatment synergistically integrated the pathways modulated by single agent treatment and promoted strong adaptive immune responses, and responses in pathways associated with cell cycle and metabolic activity. Although single agent therapy expanded intratumoral CD8⁺ T cell clones and modulated critical gene pathways, this was not sufficient for complete and long lasting tumor rejection. The findings indicate that a profound reprogramming of dysfunctional tumor infiltrating T cells by combination therapy was required for complete tumor rejection. This result is supported by a recent study showing that CD8⁺ T cells become dysfunctional at early stages of tumor development and gradually evolve into a less flexible state.

[00158] Next-generation sequencing technology has made whole-genome and transcriptome sequencing routine and provided opportunities for detection of whole genome gene expression and extraction of TCR sequences simultaneously. However, unlike conventional TCR-seq methods using targeted TCR amplicon sequencing with long reads (2x300bp), very small portion of random priming RNA-seq reads are TCR sequences and also the read length is short (usually ≤ 100 bp), which usually only cover part of V(D)J regions of TCRs. The rpsTCR pipeline was developed for assembling and extracting TCR-CDR3 sequences from random priming short RNA

sequencing reads to address this problem (**FIG. 17**). The rpsTCR took paired- and single-end short reads and mapped these reads to mouse or human genomes and transcriptomes, but not TCR gene loci and transcripts using Tophat with default parameters. Mapped reads were discarded and unmapped reads were recycled for extraction of TCR sequences. Low quality nucleotides in the unmapped reads were trimmed. Then reads with length less than 35bp were filtered out using HTQC toolkit. QC passed short reads were assembled into longer reads using iSSAKE default setting. TCRklass was used to identify CDR3 sequences with Scr (conserved residue support score) set from default 1.7 to 2. A targeted TCR-seq data from a healthy human PBMC samples was used as a positive control to evaluate whether the extra steps introduced to the pipeline result in higher false positive or false negative rates comparing to TCRklass alone. Majority of unique CDR3 sequences from TCRB (64,031) or TCRA (51,448) were detected by both rpsTCR and TCRklass. The squared correlations between rpsTCR and TCRklass were 0.9999 and 0.9365 for TCRB-CDR3 and TCRA-CDR3, respectively (**FIG. 17**). Six TCR-negative cancer or non-cancer cell lines were used as negative controls. rpsTCR didn't detected any CDR3 sequences, while TCRklass extracted a few CDR3 sequences from some of these TCR-negative cancer cell lines (**FIG. 13**). To further validate the performance of the pipeline, a healthy mouse PBMC sample was sequenced using both targeted TCR-seq and random priming RNA-seq approaches (200M, 2x100bp). Although the number of CDR3 sequences assembled from RNA-seq data was much smaller than that from the targeted TCR-seq approach, about 45% of the CDR3 sequences identified from RNA-seq data using rpsTCR were also observed among CDR3 sequences from targeted TCR-seq. Because of the technique limitation of targeted TCR-seq, it is not surprising that a fraction of the CDR3 sequences extracted from RNA-seq data were not present in the TCR-seq results. For example, the efficiency of 5' race adapter used for targeted TCR-seq is generally low and the multiply PCR tends to amplify high frequency TCRs, thus only a small portion of TCRs can be targeted. As expected, much higher percentage (~ 70%) of the CDR3 sequences identified from RNA-seq data using rpsTCR were also observed among high frequency CDR3 sequences ($\geq 0.1\%$) from targeted TCR-seq, while only about 40% extracted using TCRklass alone. Moreover, the 100bp read length was cut into 50bp segments and randomly selected 200M reads. Among the top 10 CDR3 sequences

ranked by targeted TCR-seq approach, 8 CDR3 sequences were detected by the rpsTCR, while only 3 were detected by TCRklass. The rpsTCR pipeline was then applied to extracting CDR3 sequences from the single cell RNA-seq data generated from intratumoral CD8 T cells of MC38 treated with different antibodies. The CDR3_beta and CDR3_alpha sequence detection rates were comparable to published data (**FIG. 14**) using targeted TCR-seq approach to detect TCR sequences from single cell sequencing of T cells.

- [00159] While the methods and systems have been described in connection with preferred embodiments and specific examples, it is not intended that the scope be limited to the particular embodiments set forth, as the embodiments herein are intended in all respects to be illustrative rather than restrictive.
- [00160] Unless otherwise expressly stated, it is in no way intended that any method set forth herein be construed as requiring that its steps be performed in a specific order. Accordingly, where a method claim does not actually recite an order to be followed by its steps or it is not otherwise specifically stated in the claims or descriptions that the steps are to be limited to a specific order, it is in no way intended that an order be inferred, in any respect. This holds for any possible non-express basis for interpretation, including: matters of logic with respect to arrangement of steps or operational flow; plain meaning derived from grammatical organization or punctuation; the number or type of embodiments described in the specification.
- [00161] Throughout this application, various publications are referenced. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which the methods and systems pertain.
- [00162] It will be apparent to those skilled in the art that various modifications and variations can be made without departing from the scope or spirit. Other embodiments will be apparent to those skilled in the art from consideration of the specification and practice disclosed herein. It is intended that the specification and examples be considered as exemplary only, with a true scope and spirit being indicated by the following claims.

CLAIMS

What is claimed is:

1. A method for sequencing a T cell receptor (TCR), comprising:
 - a) sequencing short reads of less than 100 base pairs of RNA obtained from a T cell;
 - b) aligning the short reads with a reference sequence, wherein the reference sequence does not contain a TCR gene sequence, thereby generating a read set comprising mapped short reads and unmapped short reads;
 - c) discarding mapped short reads from the read set;
 - d) assembling the unmapped short reads remaining in the read set into one or more long reads;
 - e) translating the one or more long reads into corresponding amino acid sequences;
 - f) fractioning TCR V region and TCR J region amino acid reference sequences into k-strings of six amino acids, aligning the k-strings with the corresponding amino acid sequences from step (e), detecting one or more conserved TCR CDR3 residues in the k-strings that map to the corresponding amino acid sequences, scoring the level of conservation detected, and selecting corresponding amino acid sequences with a conservation score above a threshold conservation score, and detecting a candidate CDR3 region amino acid sequence in the selected corresponding amino acid sequences;
 - g) identifying the nucleic acid sequence of the candidate CDR3 region amino acid sequences in the one or more long reads;
 - h) aligning the nucleic acid sequence of the one or more long reads upstream of the candidate CDR3 region nucleic acid sequence with one or more TCR V gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR V gene sequence; and
 - i) aligning the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more TCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR J gene sequence, thereby generating a TCR sequence.

2. The method of claim 1, wherein the short reads are obtained from random-priming of RNA.
3. The method of claim 1, wherein the T cell is obtained from a human or mouse.
4. The method of claim 1, wherein the reference sequence comprises a human genome, a mouse genome, a human transcriptome, or a mouse transcriptome.
5. The method of claim 1, wherein discarding mapped short reads from the read set further comprises discarding unmapped short reads from the read set that are less than 35 base pairs.
6. The method of claim 1, wherein assembling the unmapped short reads remaining in the read set into one or more long reads comprises:
aligning the one or more unmapped short reads to one or more TCR sequences from a reference database of TCR sequences; and
assembling, based on the alignment, the one or more unmapped short reads into long reads.
7. The method of claim 1 further comprising appending a TCR C region nucleic acid sequence to the TCR sequence.
8. The method of claim 1, wherein prior to sequencing the short reads of less than 100 base pairs of RNA obtained from the T cell, an immunotherapy has been administered to a subject from which the T cell is obtained.
9. The method of claim 8, wherein the immunotherapy comprises a monotherapy or a combination therapy.
10. The method of claim 9, wherein the combination therapy comprises a costimulatory agonist and a coinhibitory antagonist.
11. The method of claim 1, further comprising:

repeating steps a-i for a first plurality of T cells of a subject, wherein the T cells are collected prior to administration of a treatment;
determining a number of occurrences of unique TCR sequences present in the first plurality of T cells;
repeating steps a-i for a second plurality of T cells of the subject, wherein the T cells are collected after the administration of the treatment;
determining a number of occurrences of unique TCR sequences present in the second plurality of T cells; and
determining, based on the numbers of occurrences of unique TCR sequences present in the first plurality of T cells and the second plurality of T cell, one or more unique TCR sequences that experienced clonal expansion.

12. The method of claim 11, further comprising determining a T cell clonal expansion signature based on the one or more unique TCR sequences that experienced clonal expansion.
13. The method of claim 12, further comprising:
 - querying a database of T cell clonal expansion signatures and corresponding treatment responses using the T cell clonal expansion signature;
 - determining, based on the query, the subject's likelihood of responding to the treatment.
14. The method of claim 12, further comprising:
 - determining the subject's response to the treatment;
 - storing the T cell clonal expansion signature in a database; and
 - associating the subject's response to the treatment with the T cell clonal expansion signature in the database.
15. The method of claim 1, further comprising:
 - determining that the TCR sequence is present in a T cell clone that expands in response to a treatment;
 - producing one or more T cells containing the TCR sequence.

16. The method of claim 1, wherein sequencing short reads of less than 100 base pairs of RNA obtained from a T cell comprises bulk sequencing of short reads of less than 100 base pairs of RNA obtained from a plurality of T cells.
17. The method of claim 16, further comprising performing steps b-i for each of the plurality of T cells.
18. The method of claim 17, wherein performing steps b-i for each of the plurality of T cells comprising performing steps b-i comprises:
 - classifying at least a portion of one or more of steps b-i as a job; and
 - distributing a workload for each job across a plurality of processors in parallel.
19. An apparatus comprising:
 - one or more processors; and
 - a memory comprising processor executable instructions that, when executed by the one or more processors, cause the apparatus to:
 - a) receive sequence data comprising short reads of less than 100 base pairs of RNA obtained from a T cell;
 - b) align the short reads with a reference sequence, wherein the reference sequence does not contain a TCR gene sequence, thereby generating a read set comprising mapped short reads and unmapped short reads;
 - c) discard mapped short reads from the read set;
 - d) assemble the unmapped short reads remaining in the read set into one or more long reads;
 - e) translate the one or more long reads into corresponding amino acid sequences;
 - f) fraction TCR V region and TCR J region amino acid reference sequences into k-strings of six amino acids, aligning the k-strings with the corresponding amino acid sequences from step (e), detecting one or more conserved TCR CDR3 residues in the k-strings that map to the corresponding amino acid sequences, scoring the level of conservation detected, and selecting corresponding amino acid sequences with a conservation score above a threshold conservation score, and detecting a candidate CDR3 region amino acid sequence in the selected corresponding amino acid sequences;

- g) identify the nucleic acid sequence of the candidate CDR3 region amino acid sequences in the one or more long reads;
 - h) align the nucleic acid sequence of the one or more long reads upstream of the candidate CDR3 region nucleic acid sequence with one or more TCR V gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR V gene sequence; and
 - i) align the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more TCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR J gene sequence, thereby generating a TCR sequence.
20. A computer readable medium, having computer executable instructions embodied thereon, configured for performing a method comprising:
- a) sequencing short reads of less than 100 base pairs of RNA obtained from a T cell;
 - b) aligning the short reads with a reference sequence, wherein the reference sequence does not contain a TCR gene sequence, thereby generating a read set comprising mapped short reads and unmapped short reads;
 - c) discarding mapped short reads from the read set;
 - d) assembling the unmapped short reads remaining in the read set into one or more long reads;
 - e) translating the one or more long reads into corresponding amino acid sequences;
 - f) fractioning TCR V region and TCR J region amino acid reference sequences into k-strings of six amino acids, aligning the k-strings with the corresponding amino acid sequences from step (e), detecting one or more conserved TCR CDR3 residues in the k-strings that map to the corresponding amino acid sequences, scoring the level of conservation detected, and selecting corresponding amino acid sequences with a conservation score above a threshold conservation score, and detecting a candidate CDR3 region amino acid sequence in the selected corresponding amino acid sequences;
 - g) identifying the nucleic acid sequence of the candidate CDR3 region amino acid sequences in the one or more long reads;

- h) aligning the nucleic acid sequence of the one or more long reads upstream of the candidate CDR3 region nucleic acid sequence with one or more TCR V gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR V gene sequence; and
- i) aligning the nucleic acid sequence of the one or more long reads downstream of the candidate CDR3 region nucleic acid sequence with one or more TCR J gene reference sequences, scoring the degree of alignment, and identifying long reads above a threshold alignment score as comprising a candidate TCR J gene sequence, thereby generating a TCR sequence.