

CLAIMS

What is claimed is:

1. A method for sequencing a T cell receptor (TCR), comprising:
 - a) sequencing short reads of less than about 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 about 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. 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 about 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 about 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.
3. A computer readable medium, having computer executable instructions embodied thereon, configured for performing a method comprising:
- a) sequencing short reads of less than about 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 about 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.
4. The method, apparatus, or computer readable medium of claims 1-3, wherein the short reads are obtained from random-priming of RNA.
5. The method, apparatus, or computer readable medium of any one of claims 1-4, wherein the T cell is obtained from a human or mouse.
6. The method, apparatus, or computer readable medium of any one of claims 1-5, wherein the reference sequence comprises a human genome, a mouse genome, a human transcriptome, or a mouse transcriptome.

7. The method, apparatus, or computer readable medium of any one of claims 1-6, wherein 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.
8. The method, apparatus, or computer readable medium of any one of claims 1-7, 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.
9. The method, apparatus, or computer readable medium of any one of claims 1-8 further comprising appending a TCR C region nucleic acid sequence to the TCR sequence.
10. The method, apparatus, or computer readable medium of any one of claims 1-9 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.
11. The method, apparatus, or computer readable medium of claim 10, wherein the immunotherapy comprises a monotherapy or a combination therapy.
12. The method, apparatus, or computer readable medium of claim 11, wherein the combination therapy comprises a costimulatory agonist and a coinhibitory antagonist.
13. The method, apparatus, or computer readable medium of any one of claims 1-12, 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;

administering the treatment to the subject;

- 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.
14. The method, apparatus, or computer readable medium of claim 13, further comprising determining a T cell clonal expansion signature based on the one or more unique TCR sequences that experienced clonal expansion.
15. The method, apparatus, or computer readable medium of claim 14, 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.
16. The method, apparatus, or computer readable medium of any one of claims 14-15, 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.
17. The method, apparatus, or computer readable medium of any one of claims 1-16, 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.

18. The method, apparatus, or computer readable medium of any one of claims 1-17, 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.
19. The method, apparatus, or computer readable medium of claim 18, further comprising performing steps b-i for each of the plurality of T cells.
20. The method, apparatus, or computer readable medium of claim 19, 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.
21. A method for sequencing a B cell receptor (BCR), comprising:
- a) sequencing short reads of less than about 100 base pairs of RNA obtained from a B cell;
 - b) 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;
 - 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 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 step (e), 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;

- 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 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
- 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 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.

22. 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 about 100 base pairs of RNA obtained from a B cell;
- b) 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;
- 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 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 step (e), 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;
- 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 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
 - 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 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.
23. A computer readable medium, having computer executable instructions embodied thereon, configured for performing a method comprising:
- a) sequencing short reads of less than about 100 base pairs of RNA obtained from a B cell;
 - b) 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;
 - 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 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 step (e), 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;
 - 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 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

- 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 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.
24. The method, apparatus, or computer readable medium of claims 21-23, wherein the short reads are obtained from random-priming of RNA.
25. The method, apparatus, or computer readable medium of any one of claims 21-24, wherein the B cell is obtained from a human or mouse.
26. The method, apparatus, or computer readable medium of any one of claims 21-25, wherein the reference sequence comprises a human genome, a mouse genome, a human transcriptome, or a mouse transcriptome.
27. The method, apparatus, or computer readable medium of any one of claims 21-26, wherein 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.
28. The method, apparatus, or computer readable medium of any one of claims 21-27, 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 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.
29. The method, apparatus, or computer readable medium of any one of claims 21-28 further comprising appending a BCR C region nucleic acid sequence to the BCR sequence.
30. The method, apparatus, or computer readable medium of any one of claims 21-29 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.

31. The method, apparatus, or computer readable medium of claim 30, wherein the immunotherapy comprises a monotherapy or a combination therapy.
32. The method, apparatus, or computer readable medium of claim 31, wherein the combination therapy comprises a costimulatory agonist and a coinhibitory antagonist.
33. The method, apparatus, or computer readable medium of any one of claims 21-32, further comprising:
 - repeating steps a-i 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.
34. The method, apparatus, or computer readable medium of claim 33, further comprising determining a B cell clonal expansion signature based on the one or more unique BCR sequences that experienced clonal expansion.
35. The method, apparatus, or computer readable medium of claim 34, 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.

36. The method, apparatus, or computer readable medium of any one of claims 34-35, 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.
37. The method, apparatus, or computer readable medium of any one of claims 21-36, 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.
38. The method, apparatus, or computer readable medium of any one of claims 21-37, wherein 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.
39. The method, apparatus, or computer readable medium of claim 38, further comprising performing steps b-i for each of the plurality of B cells.
40. The method, apparatus, or computer readable medium of claim 39, wherein performing steps b-i for each of the plurality of B 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.