

## CLAIMS

What is claimed is:

1. A method for sequencing a B cell receptor (BCR), comprising:
  - a) sequencing short reads of less than 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 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.
2. The method of claim 1, wherein the short reads are obtained from random-priming of RNA.

3. The method of claim 1 or claim 2, wherein the B cell is obtained from a human or mouse.
4. The method of any one of claims 1 to 3, wherein the reference sequence comprises a human genome, a mouse genome, a human transcriptome, or a mouse transcriptome.
5. The method of any one of claims 1 to 4, 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 or that have a low sequence resolution.
6. The method of any one of claims 1 to 5, 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.
7. The method of any one of claims 1 to 6, further comprising appending a BCR C region nucleic acid sequence to the BCR sequence.
8. The method of any one of claims 1 to 7, wherein prior to sequencing the short reads of less than 100 base pairs of RNA obtained from the B cell, an immunotherapy has been administered to a subject from which the B 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 any one of claims 1 to 10, 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;  
repeating steps a-i for a second plurality of B cells of the subject, wherein the B cells are collected after 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.

12. The method of claim 11, further comprising determining a B cell clonal expansion signature based on the one or more unique BCR sequences that experienced clonal expansion.

13. The method of claim 12, 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.

14. The method of claim 12, 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.

15. The method of any one of claims 1 to 14, further comprising:  
determining that the BCR sequence is present in a B cell clone that expands in response to a treatment; and  
producing one or more B cells containing the BCR sequence.

16. The method of any one of claims 1 to 15, wherein sequencing short reads of less than 100 base pairs of RNA obtained from a B cell comprises bulk sequencing of short

reads of less than 100 base pairs of RNA obtained from a plurality of B cells.

17. The method of claim 16, further comprising performing steps b-i for each of the plurality of B cells.
18. The method of claim 17, 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.
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 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 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.

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 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 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.