

CLAIMS

1. A system for sequencing polynucleotides, comprising:
 - a sequencing system configured to generate raw sequencing data using sequencing reagents;
 - a memory comprising a reference nucleotide sequence;
 - a processor configured to control the sequencing system, and to execute instructions that perform a method comprising:
 - receiving a first nucleotide subsequence of a read from the sequencing system;
 - processing the first nucleotide subsequence to determine a first plurality of candidate locations of the read on a reference sequence using a first process;
 - determining whether the first nucleotide subsequence aligns to the reference sequence with a first confidence based on the determined candidate locations;
 - receiving a second nucleotide subsequence from the sequencing system, wherein the second nucleotide sequence comprises the first nucleotide sequence plus one or more additional nucleotides;
 - processing the second nucleotide subsequence to determine a second plurality of candidate locations of the read that align to the reference sequence using:
 - a second, simple alignment process that aligns the one or more additional nucleotides in the second nucleotide sequence to the reference sequence if the first nucleotide sequence is aligned to the reference sequence at the first confidence, or
 - repeating the first process on the second nucleotide sequence if the second nucleotide sequence is not aligned to the reference sequence at the first confidence,
 - wherein the second, simple alignment process is more computationally efficient than the first alignment process.
2. The system of claim 1, wherein the second nucleotide subsequence is processed using either the first process or the second process based on a number of mismatches or a probability of a correct match.

3. The system of claim 1, wherein the first nucleotide subsequence is one or more nucleotides in length.
4. The system of claim 1, wherein the second nucleotide subsequence is one or more nucleotides in length.
5. The system of claim 1, wherein the second process is more computationally efficient than the first process in memory usage or the number of computation operations.
6. The system of claim 1, wherein the processor is further configured to store data corresponding to at least one of the first plurality of candidate locations if the first nucleotide subsequence is aligned to the reference sequence.
7. The system of claim 6, wherein the processor is further configured to store data corresponding to at least one of the second plurality of candidate locations if the read remains aligned to the reference sequence.
8. The system of claim 1, wherein processing the second nucleotide subsequence using the second process comprises performing a simple alignment to determine a simple alignment score.
9. The system of claim 8, wherein performing the simple alignment comprises comparing the second nucleotide subsequence with corresponding sequences of the second nucleotide subsequence on the reference sequence based on the first plurality of candidate positions.
10. The system of claim 8, wherein processing the second nucleotide subsequence using the second process further comprises determining a mapping quality (MapQ) score for each of the second plurality of candidate locations of the read.
11. The system of claim 10, wherein the simple alignment score comprises the MapQ score.
12. The system of claim 1, wherein the processor is further configured to perform variant calling on the output of the first process or second process comprising at least one of the first plurality of candidate locations or at least one of the second plurality of candidate locations.

13. The system of claim 12, wherein performing the variant calling on the output of the first process or second process comprises:

performing variant calling on the output of the first process or second process using a first variant calling process or a second variant calling process, wherein the second variant calling process is more computationally efficient than the first variant calling process in variant calling of the second subsequence.

14. The system of claim 12, wherein the variant calling is performed using the output of the first process or second process based on a variant calling metric.

15. The system of claim 14, wherein the variant calling metric is determined based on a number of different base types called at a position of the reference sequence.

16. The system of claim 1, wherein processing the first nucleotide subsequence is completed before the sequencing system determines the second nucleotide subsequence during a sequencing run.

17. The system of claim 1, wherein the sequencing system implements a sequencing-by-synthesis method to determine the first subsequence.

18. A computer-implemented method for efficient sequencing of polynucleotides, comprising:

receiving a first nucleotide subsequence of a read from a sequencing system during a sequencing run;

performing a secondary analysis of the first nucleotide subsequence of the read based on a reference sequence using a first process or a second process, wherein the second process is more computationally efficient than the first process in performing the secondary analysis, and wherein the secondary analysis comprises:

comparing the first nucleotide subsequence to the reference sequence to determine a first subsequence of the reference sequence that has a high degree of similarity to the first nucleotide subsequence and to determine differences between the first nucleotide subsequence and the first subsequence of the reference sequence; and

determining if the sequencing system should generate additional nucleotide reads.

19. The method of claim 18, wherein performing the secondary analysis comprises processing the first nucleotide subsequence to determine a first plurality of candidate locations of the read that align to the reference sequence using:

the first process if the read is not aligned to the reference sequence in the previously iteration,

the second process if otherwise,

wherein the second process is more computationally efficient than the first process to determine the first plurality of candidate locations of the read.

20. The method of claim 19, wherein processing the second nucleotide subsequence using the second process comprises performing a simple alignment to determine a simple alignment score.

21. The method of claim 19, wherein results of the secondary analysis comprises output of the first process, output of the second process, or any combination thereof.

22. The method of claim 18, wherein performing the secondary analysis comprises performing variant calling of the first nucleotide subsequence, comprising:

performing variant calling on the output of the first process or the second process using a first variant calling process or a second variant calling process, wherein the second variant calling process is more computationally efficient than the first variant calling process in variant calling of the first subsequence.

23. The method of claim 22, wherein results of the secondary analysis comprises output of the first variant calling process, output of the second variant calling process, or any combination thereof.

24. The method of claim 18, further comprising providing a user with results of the secondary analysis during the sequencing run.

25. The method of claim 24, wherein the results of the secondary analysis are provided to the user at fixed intervals.

26. The method of claim 24, wherein the results of the secondary analysis are provided to the user at request of the user.

27. The method of claim 18, wherein performing the secondary analysis comprises performing the secondary analysis of the first nucleotide subsequence of the read based on results from a prior sequencing interval of the sequencing run.