

CLAIMS

What is claimed is:

1. A real-time quality control method, comprising:
generating sequence data from a biological sample using a sequencing device
conducting a sequencing run;
identifying k-mers in the sequence data that have an exact match in a hash table that is initialized with a set of k-mers comprising reference allele k-mers and alternative allele k-mers of the reference allele;
determining a distribution of the reference allele and the alternative allele in the sequence data based on a count of the exact matches; and
generating a quality metric for the biological sample based on the distribution and during the sequencing run of the biological sample.
2. The method of claim 1, comprising flagging the biological sample as contaminated based on the quality metric.
3. The method of claim 2, wherein the alternative allele is present in 5% or less of sequence reads of the sequence data in the contaminated sample.
4. The method of claim 1, comprising indicating that the biological sample passes the quality metric based on a ratio of the reference allele to the alternative allele in the sequence data being within an expected range.
5. The method of claim 1, wherein the quality metric is generated on the sequencing device.
6. The method of claim 1, wherein the alternative allele comprises a previously characterized single nucleotide polymorphism.

7. A sequencing device, comprising:
 - a substrate having loaded thereon a sequencing library prepared from a sample;
 - a computer programmed to:
 - cause the sequencing device to conduct a sequencing run to generate sequence data from sequencing library;
 - identify k-mers in the sequence data that have an exact match in a hash table that is initialized with a set of k-mers comprising reference allele k-mers and alternative allele k-mers of the reference allele;
 - determine a distribution of the reference allele and the alternative allele in the sequence data based on a count of the exact matches; and
 - generate a quality metric on the sequencing device for the biological sample based on the distribution during the sequencing run.
8. The sequencing device of claim 7, comprising a display that displays the quality metric.
9. The sequencing device of claim 7, comprising communication circuitry that communicates the generated sequence data to a cloud computing environment based on the quality metric of the biological sample being associated with passing.
10. The sequencing device of claim 9, wherein the quality metric of the biological sample is associated with a ratio of the reference allele and the alternative allele being within an expected range.
11. The sequencing device of claim 7, comprising communication circuitry that halts communication of the generated sequence data to a cloud computing environment based on the quality metric of the biological sample being associated with failing.

12. The sequencing device of claim 11, wherein the quality metric of the biological sample is associated with failing based the alternative allele being present in 5% or less of sequence reads of the sequence data.
13. A method of variant detection in a biological sample, comprising:
 - generating amplicons from a biological sample using primer pairs;
 - preparing a sequencing library from the generated amplicons;
 - generating sequence data from the sequencing library;
 - identifying sequence reads in the sequence data that start within a primer region of a primer of an individual primer pair and that are in a same direction as the primer;
 - trimming the identified sequence reads that are in the same direction as the primer to exclude sequences in the primer region; and
 - identifying a variant sequence in untrimmed sequence reads that span the primer region or that are in a different direction than the primer and at a location in the untrimmed sequence reads that correspond to or are complementary to the primer region.
14. The method of claim 13, comprising extracting RNA from the biological sample and converting the RNA to cDNA before generating the amplicons.
15. The method of claim 13, wherein the amplicons comprise overlapping portions of a reference genome.
16. The method of claim 15, wherein the reference genome is a pathogen genome.
17. The method of claim 15, wherein the reference genome is a SARS-CoV-2 genome.
18. The method of claim 16, wherein the reference genome is a human genome.

19. The method of claim 13, comprising calling the identified variant sequence based on the variant sequence being present in at least 50% of the untrimmed sequence reads at the location.

20. The method of claim 13, comprising identifying sequence reads in the sequence data that start within a reverse primer region of a reverse primer of the primer pair and that are in a same direction as the reverse primer and trimming the identified sequence reads to exclude sequences in the reverse primer region.