CLAIMS

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 ratio of a count of the exact matches to the reference allele k-mers and a count of the exact matches to the alternative allele k-mers; and

generating a sample quality metric for the biological sample based on the distribution and during the sequencing run of the biological sample, wherein the sample quality metric indicates the biological sample as passing when the ratio in the distribution is within an expected range and as failing when the ratio deviates from the expected range.

2. The method of claim 1, comprising flagging the biological sample as contaminated based on the biological sample failing the sample 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, wherein the expected range is within a tolerance of a 1:1 ratio or 1:0 ratio of the reference allele to the alternative allele.

5. The method of claim 1, wherein the sample 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 biological sample;

a computer programmed to:

cause the sequencing device to conduct a sequencing run to generate sequence data from the 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 ratio of a count of the exact matches to the reference allele k-mers and a count of the exact matches to the alternative allele k-mers; and

generate a sample quality metric on the sequencing device for the biological sample based on the distribution during the sequencing run, wherein the sample quality metric indicates the biological sample as passing when the ratio in the distribution is within an expected range and as failing when the ratio deviates from the expected range.

8. The sequencing device of claim 7, comprising a display that displays the sample 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 sample quality metric of the biological sample being associated with passing.

10. The sequencing device of claim 9, wherein the expected range is within a tolerance of a 1:1 ratio or 1:0 ratio of the reference allele to the alternative allele.

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 sample quality metric of the biological sample being associated with failing.

12. The sequencing device of claim 11, wherein the sample quality metric of the biological sample is associated with failing based on the alternative allele being present in 5% or less of sequence reads of the sequence data.