CLAIMS

1. A system comprising:

at least one processor; and

a non-transitory computer readable medium storing instructions that, when executed by the at least one processor, cause the system to:

receive, for one or more genomic coordinates of a genomic sample, a first genotype call corresponding to a first type of nucleotide reads shorter than a threshold number of nucleobases and a second genotype call corresponding to a second type of nucleotide reads satisfying the threshold number of nucleobases;

identify a first set of sequencing metrics corresponding to the first genotype call and a second set of sequencing metrics corresponding to the second genotype call;

generate, utilizing a genotype-call-integration machine-learning model to process the first set of sequencing metrics and the second set of sequencing metrics, genotype probabilities of genotype calls for the one or more genomic coordinates; and

generate an output genotype call for the one or more genomic coordinates of the genomic sample based on the genotype probabilities.

2. The system of claim 1, wherein:

the first type of nucleotide reads comprise nucleotide reads synthesized from sample library fragments that are shorter than the threshold number of nucleobases; and

the second type of nucleotide reads comprises:

assembled nucleotide reads that have been assembled from shorter nucleotide reads to form a contiguous sequence satisfying the threshold number of nucleobases;

circular consensus sequencing (CCS) reads satisfying the threshold number of nucleobases; or

nanopore long reads satisfying the threshold number of nucleobases.

3. The system of claim 1 or 2, further storing instructions that, when executed by the at least one processor, cause the system to:

generate the genotype probabilities by generating the genotype probabilities for one or more candidate single nucleotide polymorphisms (SNPs) utilizing the genotype-call-integration machine-learning model trained with SNP training data; and

generate the output genotype call indicating a presence or absence of an SNP at the one or more genomic coordinates of the genomic sample.

4. The system of any one of claims 1-3, further storing instructions that, when executed by the at least one processor, cause the system to generate the output genotype call by:

selecting the first genotype call or the second genotype call; or

generating a different genotype call differing from the first genotype call and the second genotype call.

5. The system of any one of claims 1-4, further storing instructions that, when executed by the at least one processor, cause the system to generate the genotype probabilities by:

generating a first genotype probability of the genomic sample comprising a homozygous reference genotype at the one or more genomic coordinates;

generating a second genotype probability of the genomic sample comprising a heterozygous variant genotype at the one or more genomic coordinates; and

generating a third genotype probability of the genomic sample comprising a homozygous variant genotype at the one or more genomic coordinates.

6. The system of any one of claims 1-5, wherein the first genotype call comprises a first variant call or a first reference call, and the second genotype call comprises a second variant call or a second reference call.

7. The system of any one of claims 1-6, wherein the first genotype call or the second genotype call comprises a null-data indicator.

8. The system of any one of claims 1-7, further storing instructions that, when executed by the at least one processor, cause the system to:

modify a genotype metric, a base-call-quality metric, a genotype quality metric, a genotype probability metric, a genotype-likelihood metric, or a PHRED-scaled-genotype-likelihood metric based on the genotype probabilities; and

generate a variant call file that includes the modified genotype metric, the modified base-call-quality metric, the modified genotype quality metric, the modified genotype probability metric, the modified genotype-likelihood metric, or the modified PHRED-scaled-genotype-likelihood metric.

9. The system of any one of claims 1-8, further storing instructions that, when executed by the at least one processor, cause the system to generate the output genotype call by selecting the first genotype call instead of the second genotype call by:

selecting a homozygous reference genotype call from the first genotype call instead of a heterozygous variant genotype call or a homozygous variant genotype call from the second genotype call;

selecting the heterozygous variant genotype call from the first genotype call instead of the homozygous reference genotype call or the homozygous variant genotype call from the second genotype call; or

selecting the homozygous variant genotype call from the first genotype call instead of the heterozygous variant genotype call or the homozygous reference genotype call from the second genotype call.

10. The system of any one of claims 1-9, further storing instructions that, when executed by the at least one processor, cause the system to generate the output genotype call by selecting the second genotype call instead of the first genotype call by:

selecting a homozygous reference genotype call from the second genotype call instead of a heterozygous variant genotype call or a homozygous variant genotype call from the first genotype call;

selecting the heterozygous variant genotype call from the second genotype call instead of the homozygous reference genotype call or the homozygous variant genotype call from the first genotype call; or

selecting the homozygous variant genotype call from the second genotype call instead of the heterozygous variant genotype call or the homozygous reference genotype call from the first genotype call.

11. A non-transitory computer readable medium storing instructions that, when executed by at least one processor, cause a system to:

receive, for one or more genomic coordinates of a genomic sample, a first genotype call corresponding to a first type of nucleotide reads shorter than a threshold number of nucleobases and a second genotype call corresponding to a second type of nucleotide reads satisfying the threshold number of nucleobases;

identify a first set of sequencing metrics corresponding to the first genotype call and a second set of sequencing metrics corresponding to the second genotype call;

generate, utilizing a genotype-call-integration machine-learning model to process the first set of sequencing metrics and the second set of sequencing metrics, genotype probabilities of genotype calls for the one or more genomic coordinates; and

generate an output genotype call for the one or more genomic coordinates of the genomic sample based on the genotype probabilities.

12. The non-transitory computer readable medium of claim 11, further storing instructions that, when executed by the at least one processor, cause the system to identify the first set of sequencing metrics corresponding to the first genotype call and the second set of sequencing metrics corresponding to the second genotype call by identifying a shared set of sequencing metrics associated with both the first genotype call and the second genotype call.

13. The non-transitory computer readable medium of claim 11 or 12, further storing instructions that, when executed by the at least one processor, cause the system to identify the first set of sequencing metrics corresponding to the first genotype call or the second set of sequencing metrics corresponding to the second genotype call by determining one or more of read-based sequencing metrics, call-model-generated sequencing metrics, externally sourced sequencing metrics, or second-read-type sequencing metrics associated with the second genotype call corresponding to the second type of nucleotide reads.

14. The non-transitory computer readable medium of any one of claims 11-13, further storing instructions that, when executed by the at least one processor, cause the system to identify the first set of sequencing metrics corresponding to the first genotype call or the second set of sequencing metrics corresponding to the second genotype call by identifying read-based sequencing metrics comprising one or more of:

an allele frequency corresponding to an allele for the first genotype call, an allele for the second genotype call, or a different allele for an alternative genotype call differing from the first and second genotype calls;

a coverage depth of the first type of nucleotide reads corresponding to the first genotype call or the second type of nucleotide reads corresponding to the second genotype call;

an average coverage depth of the first type of nucleotide reads corresponding to the first genotype call or the second type of nucleotide reads corresponding to the second genotype call;

a mapping-quality metric for the first type of nucleotide reads corresponding to the first genotype call or the second type of nucleotide reads corresponding to the second genotype call; or

a nucleobase composition of one or more nucleotide reads from the first type of nucleotide reads or the second type of nucleotide reads.

15. The non-transitory computer readable medium of any one of claims 11-14, further storing instructions that, when executed by the at least one processor, cause the system to identify the first set of sequencing metrics corresponding to the first genotype call or the second set of sequencing metrics corresponding to the second genotype call by identifying call-model-generated sequencing metrics comprising one or more of: a genotype metric, a base-call-quality metric, a genotype quality metric, a genotype probability metric, or a PHRED-scaled-likelihood metric for the first genotype call determined from the first type of nucleotide reads or the second genotype call determined from the second type of nucleotide reads.

16. The non-transitory computer readable medium of any one of claims 11-15, further storing instructions that, when executed by the at least one processor, cause the system to identify the first set of sequencing metrics corresponding to the first genotype call or the second

set of sequencing metrics corresponding to the second genotype call by identifying externally sourced sequencing metrics comprising one or more of:

a mappability metric indicating a degree of difficulty with which a nucleotide read is mapped to the one or more genomic coordinates within a reference genome;

a guanine-cytosine-content metric indicating a count of guanine-cytosine content corresponding to the one or more genomic coordinates within the reference genome;

a confidence classification or confidence score indicating a degree to which nucleobases at the one or more genomic coordinates can be accurately determined;

a repeat classification indicating a category of repetitive genomic region for the one or more genomic coordinates;

an indicator that the one or more genomic coordinates are part of a cytosine quadruplex (C-quadruplex) within the reference genome;

an indicator that the one or more genomic coordinates are part of a guanine quadruplex (G-quadruplex) within the reference genome; or

an indicator that the one or more genomic coordinates are part of a homopolymer within the reference genome.

17. A computer-implemented method comprising:

receiving, for one or more genomic coordinates of a genomic sample, a first genotype call corresponding to a first type of nucleotide reads shorter than a threshold number of nucleobases and a second genotype call corresponding to a second type of nucleotide reads satisfying the threshold number of nucleobases;

identifying a first set of sequencing metrics corresponding to the first genotype call and a second set of sequencing metrics corresponding to the second genotype call;

generating, utilizing a genotype-call-integration machine-learning model to process the first set of sequencing metrics and the second set of sequencing metrics, genotype probabilities of genotype calls for the one or more genomic coordinates; and

generating an output genotype call for the one or more genomic coordinates of the genomic sample based on the genotype probabilities.

18. The computer-implemented method of claim 17, wherein:

the first type of nucleotide reads comprise nucleotide reads synthesized from sample library fragments that are shorter than the threshold number of nucleobases; and

the second type of nucleotide reads comprises:

assembled nucleotide reads that have been assembled from shorter nucleotide reads to form a contiguous sequence satisfying the threshold number of nucleobases;

circular consensus sequencing (CCS) reads satisfying the threshold number of nucleobases; or

nanopore long reads satisfying the threshold number of nucleobases.

19. The computer-implemented method of claim 17 or 18, further comprising:

receive the first genotype call by receiving the first genotype call as part of a first variant call file based on the first type of nucleotide reads;

receive the second genotype call by receiving the second genotype call as part of a second variant call file based on the second type of nucleotide reads; and

generate a merged variant call file comprising the first genotype call or the second genotype call.

20. The computer-implemented method of any one of claims 17-19, further comprising:

determine that the first genotype call comprises a first alternate nucleobase that differs from a second alternate nucleobase of the second genotype call;

generate, utilizing the genotype-call-integration machine-learning model to process the first set of sequencing metrics and the second set of sequencing metrics, a first pipeline-accuracy likelihood of the first genotype call being more accurate than the second genotype call and a second pipeline-accuracy likelihood of the second genotype call being more accurate than the first genotype call; and

generate the output genotype call by selecting the first genotype call or the second genotype call for the one or more genomic coordinates of the genomic sample based on the first pipeline-accuracy likelihood and the second pipeline-accuracy likelihood.