CLAIMS

1. A computer-implemented method comprising:

accessing, for a sample nucleotide sequence, one or more sequencing data files comprising a genotype call at a genomic coordinate;

extracting, from the one or more sequencing data files, sequencing metrics for the genotype call;

generating, utilizing a call-recalibration-machine-learning model and based on the sequencing metrics, one or more variant-call classifications indicating an accuracy of the genotype call within the one or more sequencing data files; and

generating, based on the one or more variant-call classifications, a recalibrated sequencing data file comprising an updated genotype call at the genomic coordinate for the sample nucleotide sequence.

2. The computer-implemented method of claim 1, further comprising:

accessing an alignment data file of the one or more sequencing data files, the alignment data file comprising nucleotide reads corresponding to the genomic coordinate for the sample nucleotide sequence; and

extracting, from the alignment data file, one or more read-based sequencing metrics of the sequencing metrics, the one or more read-based sequencing metrics corresponding to the nucleotide reads.

3. The computer-implemented method of claim 1 or 2, further comprising generating the one or more variant-call classifications without utilizing a call-generation model to contemporaneously generate the genotype call.

4. The computer-implemented method of any one of claims 1-3, wherein extracting the sequencing metrics comprises extracting one or more read-based sequencing metrics or call-model-generated sequencing metrics for the genotype call from a genotype-call data file of the one or more sequencing data files, the genotype-call data file comprising the genotype call.

5. The computer-implemented method of claim 4, wherein accessing the genotype-call data file comprises accessing a variant call format (VCF) or a genomic variant call format (gVCF) file comprising variant and non-variant calls.

6. The computer-implemented method of claim 4, wherein the genotype-call data file comprising the genotype call was generated on a computing device executing a hardware accelerator and the recalibrated sequencing data file is generated utilizing a general-purpose processing unit.

7. The computer-implemented method of claim 6, wherein the hardware accelerator comprises a field-programmable gate array (FPGA) or an application specific integrated circuit (ASIC) and the recalibrated sequencing data file is generated utilizing one or more of a central processing unit (CPU) or a graphical processing unit (GPU).

8. The computer-implemented method of claim 4, further comprising:

accessing an additional genotype-call data file generated by a different version of a call-generation model than a version of the call-generation model that generated the genotype-call data file;

extracting, from the additional genotype-call data file, additional sequencing metrics for an additional genotype call at a genomic coordinate for an additional sample nucleotide sequence;

generating, utilizing the call-recalibration-machine-learning model and based on the additional sequencing metrics for the additional genotype call, one or more additional variant-call classifications indicating an accuracy of the additional genotype call within the additional genotype-call data file; and

generating, based on the one or more additional variant-call classifications, an additional recalibrated sequencing data file comprising an updated additional genotype call at the genomic coordinate for the additional sample nucleotide sequence.

9. The computer-implemented method of any one of claims 1-8, further comprising generating the one or more variant-call classifications generating one or more of a false-positive probability that the genotype call is a false positive, a genotype-error probability that a genotype for the genotype call is incorrect, or a true-positive probability that the genotype call is a true positive.

10. The computer-implemented method of any one of claims 1-9, wherein determining the updated genotype call comprises:

identifying the genomic coordinate as a multiallelic genomic coordinate;

generating, utilizing the call-recalibration-machine-learning model, the one or more variant-call classifications comprising one or more of a reference probability that the genotype call

comprises a homozygous reference genotype at the multiallelic genomic coordinate, a zygosity-error probability that the genotype call comprises a genotype-zygosity error at the multiallelic genomic coordinate, or a true-positive variant probability that the genotype call constitutes a true positive variant at the multiallelic genomic coordinate; and

determining the updated genotype call at the multiallelic genomic coordinate based on one or more of the reference probability, the zygosity-error probability, or the true-positive variant probability.

11. The computer-implemented method of any one of claims 1-10, further comprising:

modifying, based on the one or more variant-call classifications, one or more of a base-call-quality metric, a genotype-probability metric, a genotype metric, a genotype-likelihood metric, or a genotype-quality metric for the genotype call; and

generating the recalibrated sequencing data file comprising the modified base-call-quality metric, the modified genotype-probability metric, the modified genotype metric, the modified genotype-likelihood metric, or the modified genotype-quality metric.

12. The computer-implemented method of any one of claims 1-11, wherein generating, as part of the recalibrated sequencing data file, the updated genotype call at a biallelic genomic coordinate for the sample nucleotide sequence comprises:

determining a homozygous-reference genotype call at the genomic coordinate instead of a heterozygous-variant genotype call or a homozygous-variant genotype call reported in the one or more sequencing data files;

determining the heterozygous-variant genotype call at the genomic coordinate instead of the homozygous-reference genotype call or the homozygous-variant genotype call reported in the one or more sequencing data files; or

determining the homozygous-variant genotype call at the genomic coordinate instead of the heterozygous-variant genotype call or the homozygous-reference genotype call reported in the one or more sequencing data files.

13. The computer-implemented method of any one of claims 1-12, further comprising:

extracting, from the one or more sequencing data files, sequencing metrics for an additional genotype call at an additional genomic coordinate for the sample nucleotide sequence;

generating, utilizing the call-recalibration-machine-learning model and based on the sequencing metrics for the additional genotype call, one or more additional variant-call

classifications indicating an accuracy of the additional genotype call within the one or more sequencing data files;

modifying, based on the one or more additional variant-call classifications, a base-call-quality metric for the additional genotype call to generate a modified base-call-quality metric that falls below a base-call-quality threshold; and

annotating the additional genotype call to indicate the modified base-call-quality metric falls below the base-call-quality threshold.

14. A system comprising at least one processor and a non-transitory computer readable medium comprising instructions that, when executed by the at least one processor, cause the system to perform a method according to any of claims 1-13.

15. A non-transitory computer-readable medium storing instructions that, when executed by at least one processor, cause a system to perform a method according to any of claims 1-13.