CLAIMS

What is claimed is:

Claim 1. A computer-implemented method of variant calling, comprising:

accessing a data file comprising sequencing data output by a sequencer instrument based on processing a genomic sample;

identifying, within the data file, a group of reads that span a candidate variant at a target base position;

processing data associated with the group of reads within the data file through a neural network for evaluation of the candidate variant at the target base position;and

generating, based on the neural network processing the data associated with the group of reads, at least one classification scores indicating a likelihood that the candidate variant at the target base position is a variant.

Claim 2. The computer-implemented method according to claim 1, wherein processing data associated with the group of reads within the data file through a neural network comprises providing, to the neural network, an array of input features generated from the data associated with the group of reads.

Claim 3. The computer-implemented method according to claim 2, wherein the array of input features comprises one or more bases at the target base position, one or more bases flanking the target base position, or one or more corresponding base features for bases within the group of reads.

Claim 4. The computer-implemented method according to claim 3, wherein the one or more corresponding base features for bases within the group of reads comprise one or more of:

a corresponding reference base in a reference read corresponding to a reference genome;
a base call accuracy score of calling a base in a read of the group of reads;
a strandedness of calling the base;
an insertion count of changes adjoining a position of the base;
a deletion flag at the position of the base; or
a mapping quality score of aligning a corresponding read that contains the base to the reference read.

Claim 5. The computer-implemented method according to claims 2- 4, wherein each feature in the array of input features corresponds to a base in a read from the group of reads and each feature comprises a plurality of dimensions.

Claim 6. The computer-implemented method according to any one of claims 1 to 5, wherein the neural network comprises a deep neural network.

Claim 7. The computer-implemented method according to any one of claims 1 to 6, wherein processing data associated with the group of reads comprises processing the data associated with the group of reads through a one or more of an input layer, a convolution layer, a batch normalization layer, a max pooling layer, or a flattening layer.

Claim 8. The computer-implemented method according to any one of claims 1 to 7, wherein each of the

processing data associated with the group of reads comprises:

processing the data associated with the group of reads through one or more layers of the neural network to generate output features;

concatenating the output features with per-variant characterization data corresponding to the candidate variant at the target base position to generate concatenated features; and

generating the at least one classification score based on an analysis of the concatenated features by one or more layers of the neural network.

Claim 9. The computer-implemented method according to claim 8, wherein the per-variant characterization data corresponding to the candidate variant comprise empirical variant score (EVS) features for the candidate variant.

Claim 10. The computer-implemented method according to claim 8 or 9, wherein generating the at least one classification score based on the analysis of the concatenated features comprises utilizing one or more fully connected layers and a classification layer of the neural network.

Claim 11. The computer-implemented method of any preceding claim, wherein data file comprises a text file or a Binary Alignment Map (BAM) file.

Claim 12. The computer-implemented method of any preceding claim, wherein generating the at least one classification score comprises generating a first probability that the candidate variant is a true variant and a second probability that the candidate variant is a false variant.

Claim 13. The computer-implemented method of any preceding claim, wherein generating the at least one classification score comprises generating one or more of a first score that the candidate variant is a homozygous variant, a second score that the candidate variant is a heterozygous variant, a third score that the candidate variant is a non-variant, or a fourth score that the candidate variant is a complex-variant.

Claim 14. A non-transitory computer readable storage medium storing computer readable instructions configured to cause a processor to perform a computer-implemented method according to any one of claims 1-13.

Claim 15. A system including one or more processors coupled to memory, the memory loaded with computer instructions to call variants, the computer instructions, when executed on the processors, implement actions comprising a computer-implemented method according to any one of claims 1-13.