

CLAIMS

1. An artificial intelligence-based method of base calling, the method including:
 - training a first base caller by using a first set of cluster images as training data,
 - wherein the first set of cluster images are annotated with first ground truth data that uses discrete valued labels to identify a correct base call;
 - evaluating a second set of cluster images as inference data by applying the trained first base caller on the second set of cluster images and generating base call predictions,
 - wherein the base call predictions are represented by continuous valued weights that identify a predicted base call;
 - training a second base caller using the second set of cluster images as training data,
 - wherein the second set of cluster images are annotated with second ground truth data that identifies a correct base call based on
 - (i) the discrete valued labels, and
 - (ii) the continuous valued weights;
 - wherein the second base caller has fewer processing modules and parameters than the first base caller; and
 - evaluating a third set of cluster images as inference data by applying the trained second base caller on the third set of cluster images and generating base call predictions.
2. The artificial intelligence-based method of claim 1, wherein the discrete valued labels are one-hot encoded with a one-value for a correct base and zero-values for incorrect bases.
3. The artificial intelligence-based method of claim 2, wherein the discrete valued labels have a near-one-value for the correct base and near-zero-values for the incorrect bases.
4. The artificial intelligence-based method of claim 1, wherein the continuous valued weights are part of a probability distribution for a correct base being Adenine (A), Cytosine (C), Thymine (T), and Guanine (G).
5. The artificial intelligence-based method of claim 1, wherein one of the processing modules is neural network layers.
6. The artificial intelligence-based method of claim 5, wherein one of the parameters is interconnections between the neural network layers.
7. The artificial intelligence-based method of claim 1, wherein one of the processing modules is neural network filters.
8. The artificial intelligence-based method of claim 1, wherein one of the processing modules is neural network kernels.

9. The artificial intelligence-based method of claim 1, wherein one of the parameters is multiplication and addition operations.
10. The artificial intelligence-based method of claim 1, further including:
training the second base caller using the second set of cluster images as training data,
wherein the second set of cluster images are annotated with the second ground truth data that identifies the correct base call based on
(i) the continuous valued weights.
11. The artificial intelligence-based method of claim 1, wherein a cluster image depicts intensity emissions of clusters, and
wherein the intensity emissions are captured during a sequencing cycle of a sequencing run.
12. The artificial intelligence-based method of claim 11, wherein the cluster image further depicts intensity emissions of background surrounding the clusters.
13. The artificial intelligence-based method of claim 1, wherein the first, second, and third sets of cluster images share one or more common cluster images.
14. The artificial intelligence-based method of claim 1, further including:
training an ensemble of the first base caller by using the first set of cluster images as training data,
wherein the first set of cluster images are annotated with the first ground truth data that uses the discrete valued labels to identify the correct base call, and
wherein the ensemble comprises two or more instances of the first base caller;
evaluating the second set of cluster images as inference data by applying the trained first base caller on the second set of cluster images and generating the base call predictions,
wherein the base call predictions are represented by the continuous valued weights that identify the predicted base call;
training the second base caller using the second set of cluster images as training data,
wherein the second set of cluster images are annotated with the second ground truth data that identifies the correct base call based on
(i) the discrete valued labels, and
(ii) the continuous valued weights;
wherein the second base caller has fewer processing modules and parameters than the ensemble of the first base caller; and
evaluating the third set of cluster images as inference data by applying the trained second

base caller on the third set of cluster images and generating the base call predictions.

15. The artificial intelligence-based method of claim 1, further including:

implementing the trained second base caller on one or more parallel processors of a sequencing instrument for real-time base calling.

16. A system for artificial intelligence-based base calling, comprising:

a base caller trained on cluster images that are annotated with ground truth data that identifies a correct base call based on

(i) discrete valued labels of ground truth data used to train another base caller, and

(ii) continuous valued weights of base call predictions generated by the another base caller for the cluster images during inference;

wherein the base caller has fewer processing modules and parameters than the another base caller; and

wherein the base caller is configured to evaluate additional cluster images and generate, for the additional cluster images, base call predictions.

17. The system of claim 16, wherein the discrete valued labels are one-hot encoded with a one-value for a correct base and zero-values for incorrect bases.

18. The system of claim 16, wherein the continuous valued weights are part of a probability distribution for a correct base being Adenine (A), Cytosine (C), Thymine (T), and Guanine

(G). 19. The system of claim 16, wherein:

one of the processing modules is neural network layers,

one of the parameters is interconnections between the neural network layers,

one of the processing modules is neural network filters,

one of the processing modules is neural network kernels, and

one of the parameters is multiplication and addition operations.

20. A system for artificial intelligence-based base calling, comprising:

a first base caller trained on cluster images that are annotated with ground truth data that identifies a correct base call based on base call predictions generated by a second base caller.