CLAIMS

1. A computer-implemented method of progressively training a base caller, including:

iteratively initially training a base caller with analyte of a single-oligo base sequence with a single known oligo base sequence, and generating labelled training data using the initially trained base caller, the base caller having an initial neural network configuration comprising a number of layers and parameters;

(i) further training the base caller with analyte comprising multi-oligo base sequences comprising at least two known oligo base sequences that differ from one another by a threshold edit distance between respective nucleotide bases thereof, and generating labelled training data using the further trained base caller; and

iteratively further training the base caller by repeating step (i), while, during at least one iteration, increasing a complexity of the initial neural network configuration of the base caller by increasing the number of layers and parameters relative to the initial neural network configuration to adjust the initial neural network configuration for the multi-oligo base sequences, wherein labelled training data generated during an iteration is used to train the base caller during an immediate subsequent iteration.

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

during at least one iteration of further training the base caller with the analyte comprising multi-oligo base sequences:

increasing, within the analyte, a number of unique oligo base sequences of the multi-oligo base sequences; and

further increasing the number of layers and parameters to further adjust the initial neural network configuration for the increased number of unique oligo base sequences.

3. The computer-implemented method of claim 1 or 2, wherein iteratively initially training the base caller with the analyte comprising the single-oligo base sequence comprises:

during a first iteration of iteratively initially training the base caller:

populating the single-oligo base sequence into a plurality of clusters of a flow cell;

generating a plurality of sequence signals corresponding to the plurality of clusters, each sequence signal of the plurality of sequence signals representative of base sequences loaded in a corresponding cluster of the plurality of clusters;

predicting, based on each sequence signal of the plurality of sequence signals, corresponding base calls for the single known oligo base sequence, to thereby generate a plurality of predicted base calls;

generating, for each sequence signal of the plurality of sequence signals, a corresponding error signal, based on comparing (i) corresponding predicted base calls and (ii) bases of the single known oligo base sequence, thereby generating a plurality of error signals corresponding to the plurality of sequence signals; and

initially training the base caller during the first iteration, based on the plurality of error signals.

4. The computer-implemented method of claim 3, wherein iteratively initially training the base caller with the analyte comprising the single-oligo base sequence further comprises:

during a second iteration of iteratively initially training the base caller that occurs after the first iteration of iteratively initially training the base caller:

using the base caller that has been partially trained during the first iteration, predicting, based on each sequence signal of the plurality of sequence signals, corresponding further base calls for the single known oligo base sequence, to thereby generate a plurality of further predicted base calls;

generating, for each sequence signal of the plurality of sequence signals, a corresponding further error signal, based on comparing (i) corresponding further predicted base calls and (ii) bases of the single known oligo base sequence, thereby generating a plurality of further error signals corresponding to the plurality of sequence signals; and

further initially training the base caller during the second iteration, based on the plurality of further error signals.

5. The computer-implemented method of claim 4, wherein:

the plurality of sequence signals corresponding to the plurality of clusters, which are generated during the first iteration of iteratively initially training the base caller, is reused for the second iteration of iteratively initially training the base caller.

6. The computer-implemented method of any of claims 4-5, wherein comparing (i) corresponding predicted base calls and (ii) bases of the single known oligo base sequence comprises:

for a first predicted base calls, (i) comparing a first base of the first predicted base calls with a first base of the single known oligo base sequence and (ii) comparing a second base of the

first predicted base calls and a second base of the single known oligo base sequence, to generate a corresponding first error signal.

7. The computer-implemented method of any of claims 1-6, wherein iteratively further training the base caller comprises:

further training the base caller for N1 iterations with analyte comprising two known unique oligo base sequences; and

further training the base caller for N2 iterations with analyte comprising three known unique oligo base sequences,

wherein the N1 iterations are performed prior to the N2 iterations.

8. A system comprising:

at least one processor; and

a non-transitory computer readable storage medium impressed with computer program instructions that, when executed on the at least one processor, implement actions comprising:

iteratively initially training a base caller with analyte of a single-oligo base sequence with a single known oligo base sequence, and generating labelled training data using the initially trained base caller, the base caller having an initial neural network configuration comprising a number of layers and parameters;

(i) further training the base caller with analyte comprising multi-oligo base sequences comprising at least two known oligo base sequences that differ from one another by a threshold edit distance between respective nucleotide bases thereof, and generating labelled training data using the further trained base caller; and

iteratively further training the base caller by repeating step (i), while, during at least one iteration, increasing a complexity of the initial neural network configuration of the base caller by increasing the number of layers and parameters relative to the initial neural network configuration to adjust the initial neural network configuration for the multi-oligo base sequences, wherein labelled training data generated during an iteration is used to train the base caller during an immediate subsequent iteration.

9. The system of claim 8, wherein while iteratively initially training the base caller with the analyte comprising the single-oligo base sequence, a first neural network configuration is loaded within the base caller, and wherein iteratively further training the base caller comprises:

further training the base caller for N1 iterations with analyte comprising two known unique oligo base sequences, such that

(i) for a first subset of the N1 iterations, a second neural network configuration is loaded within the base caller, and

(ii) for a second subset of the N1 iterations occurring after the first subset of the N1 iterations, a third neural network configuration is loaded within the base caller, wherein the first, second, and third neural network configurations are different from each other.

10. The system of claim 9, wherein the second neural network configuration is more complex than the first neural network configuration, and wherein the third neural network configuration is more complex than the second neural network configuration.

11. The system of claim 9 or 10, wherein the second neural network configuration has a greater number of layers, weights, or parameters than the first neural network configuration.

12. The system of any of claims 9-11, wherein the third neural network configuration has a greater number of layers, weights, or parameters than the second neural network configuration.

13. The system of any of claims 9-12, wherein further training the base caller for the N1 iterations with the analyte comprising two known unique oligo base sequences comprises, for one iteration of the N1 iterations:

populating (i) a first plurality of clusters of a flow cell with a first known oligo base sequence of the two known unique oligo base sequences and (ii) a second plurality of clusters of the flow cell with a second known oligo base sequence of the two known unique oligo base sequences;

predicting, for each cluster of the first and second pluralities of clusters, corresponding base calls, such that a plurality of predicted base calls are generated;

mapping (i) a first predicted base call of the plurality of predicted base calls to the first known oligo base sequence and (ii) a second predicted base call of the plurality of predicted base calls to the second known oligo base sequence, while refraining from mapping a third predicted base call of the plurality of predicted base calls to any of the first or second known oligo base sequences;

generating (i) a first error signal, based on comparing the first predicted base call to the first known oligo base sequence, and (ii) a second error signal, based on comparing the second predicted base call to the second known oligo base sequence; and

further training the base caller, based on the first and second error signals.

14. The system of claim 13, wherein mapping the first predicted base call to the first known oligo base sequence of the two known unique oligo base sequences comprises:

comparing each base of the first predicted base call to corresponding bases of the first and second known oligo base sequences;

determining that the first predicted base call has at least a threshold number of similarity of bases with the first known oligo base sequence, and has less than the threshold number of similarity of bases with the second known oligo base sequence; and

based on determining that the first predicted base call has at least the threshold number of similarity of bases with the first known oligo base sequence, mapping the first predicted base call to the first known oligo base sequence.

15. The system of claim 13 or 14, wherein refraining from mapping the third predicted base call to any of the first or second known oligo base sequences comprises:

comparing each base of the first predicted base call to corresponding base of the first and second known oligo base sequences;

determining that the first predicted base call has less than a threshold number of similarity of bases with each of the first and second known oligo base sequences; and

based on determining that the first predicted base call has less than the threshold number of similarity of bases with each of the first and second known oligo base sequences, refraining from mapping the third predicted base call to any of the first or second known oligo base sequences.

16. The system of any of claims 13-15, wherein refraining from mapping the third predicted base call to any of the first or second known oligo base sequences comprises:

comparing each base of the first predicted base call to corresponding base of the first and second known oligo base sequences;

determining that the first predicted base call has more than a threshold number of similarity of bases with each of the first and second known oligo base sequences; and

based on determining that the first predicted base call has more than the threshold number of similarity of bases with each of the first and second known oligo base sequences, refraining from mapping the third predicted base call to any of the first or second known oligo base sequences.

17. The system of any of claims 13-16, wherein generating labelled training data using the further trained base caller for the one iteration of the N1 iterations comprises:

subsequent to further training the base caller during the one iteration of the N1 iterations, re-predicting, for each cluster of the first and second pluralities of clusters, corresponding base calls, such that an additional plurality of predicted base calls are generated;

remapping (i) a first subset of the additional plurality of predicted base calls to the first known oligo base sequence and (ii) a second subset of the additional plurality of predicted base

calls to the second known oligo base sequence, while refraining from mapping a third subset of the additional plurality of predicted base calls to any of the first or second known oligo base sequences; and

generating labelled training data based on the remapping, such that the labelled training data includes (i) the first subset of the additional plurality of predicted base calls, with the first known oligo base sequence forming ground truth data for the first subset of the additional plurality of predicted base calls, and (ii) the second subset of the additional plurality of predicted base calls, with the second known oligo base sequence forming ground truth data for the second subset of the additional plurality of predicted base calls.

18. The system of claim 17, wherein:

the labelled training data generated during the one iteration of the N1 iterations is used to train the base caller during an immediate subsequent iteration of the N1 iterations; and

a neural network configuration of the base caller is unchanged during the one iteration of the N1 iterations and the immediate subsequent iteration of the N1 iterations; or

a neural network configuration of the base caller during the immediate subsequent iteration of the N1 iterations is different from, and more complex than, a neural network configuration of the base caller during the one iteration of the N1 iterations.

19. A non-transitory computer readable storage medium impressed with computer program instructions to progressively train a base caller, the computer program instructions, when executed on a processor, implement actions comprising:

iteratively initially training a base caller with analyte of a single-oligo base sequence with a single known oligo base sequence, and generating labelled training data using the initially trained base caller, the base caller having an initial neural network configuration comprising a number of layers and parameters;

(i) further training the base caller with analyte comprising multi-oligo base sequences comprising at least two known oligo base sequences that differ from one another by a threshold edit distance between respective nucleotide bases thereof, and generating labelled training data using the further trained base caller; and

iteratively further training the base caller by repeating step (i), while, during at least one iteration, increasing a complexity of the initial neural network configuration of the base caller by increasing the number of layers and parameters relative to the initial neural network configuration to adjust the initial neural network configuration for the multi-oligo base sequences, wherein labelled training data generated during an iteration is used to train the base caller during an immediate subsequent iteration.

20. The non-transitory computer readable storage medium of claim 19, further impressed with computer program instructions that, when executed on the processor, implement actions comprising iteratively further training the base caller by:

prior to further training the base caller with analyte comprising multi-oligo base sequences, and in response to increasing the number of layers and parameters relative to the initial neural network configuration to adjust the initial neural network configuration for the multi-oligo base sequences, further training the base caller with labelled training data generated during a final iteration of iteratively initially training the base caller.