

CLAIMS

1. An artificial intelligence-based method of base calling index sequences, the method including:
 - accessing index images generated for the index sequences during index sequencing cycles of a sequencing run, wherein the index images depict intensity emissions generated as a result of nucleotide incorporation in the index sequences during the sequencing run;
 - preprocessing the index images using a normalization function that produces a normalized version of an index image from a current index sequencing cycle based on
 - (i) intensity values of index images from one or more preceding index sequencing cycles,
 - (ii) intensity values of index images from one or more succeeding index sequencing cycles, and
 - (iii) intensity values of index images from the current index sequencing cycle; and
 - processing normalized versions of the index images through a neural network-based base caller and generating a base call for each of the index sequencing cycles, thereby producing index reads for the index sequences.
2. The artificial intelligence-based method of claim 1, wherein the normalization function calculates:
 - a lower percentile of (i) the intensity values of the index images from the one or more preceding index sequencing cycles, (ii) the intensity values of the index images from the one or more succeeding index sequencing cycles, and (iii) the intensity values of the index images from the current index sequencing cycle, and
 - an upper percentile of (i) the intensity values of the index images from the one or more preceding index sequencing cycles, (ii) the intensity values of the index images from the one or more succeeding index sequencing cycles, and (iii) the intensity values of the index images from the current index sequencing cycle, such that,
 - in the normalized version of the index image,
 - a first percentage of normalized intensity values are below the lower percentile,
 - a second percentage of the normalized intensity values are above the upper percentile, and
 - a third percentage of the normalized intensity values are between the lower and upper percentiles.

3. The artificial intelligence-based method of claim 1, wherein,
 taken together, nucleotides depicted by the index images from the current, preceding, and succeeding index sequencing cycles
 are cumulatively more diverse than
 nucleotides depicted only by the index images from the current index sequencing cycle.
4. The artificial intelligence-based method of claim 3, wherein at least one index image in the index images from the preceding and succeeding index sequencing cycles depicts one or more nucleotides in a detectable signal state.
5. The artificial intelligence-based method of claim 3, wherein the nucleotides depicted by the index images from the current index sequencing cycle are low-complexity patterns in which some of four bases A, C, T, and G are represented at a frequency of less than 15%, 10%, or 5% of all the nucleotides.
6. The artificial intelligence-based method of claim 5, wherein, taken together, the nucleotides depicted by the index images from the current, preceding, and succeeding index sequencing cycles cumulatively form high-complexity patterns in which each of the four bases A, C, T, and G are represented at a frequency of at least 20%, 25%, or 30% of all the nucleotides.
7. The artificial intelligence-based method of claim 1, further including:
 preprocessing the index images using the normalization function during training of the neural network-based base caller as well as during inference.
8. The artificial intelligence-based method of claim 1, further including:
 preprocessing the index images using an augmentation function that produces an augmented version of an index image by multiplying intensity values of the index image with a scaling factor and adding an offset value to the multiplication's result; and
 processing augmented versions of the index images through the neural network-based base caller and generating a base call for each of the index sequencing cycles, thereby producing index reads for the index sequences.
9. The artificial intelligence-based method of claim 8, further including:
 preprocessing the index images using the augmentation function only during the training of the neural network-based base caller and not during the inference.

10. The artificial intelligence-based method of claim 1, further including:
preprocessing the index images using the normalization function that produces the normalized version of the index image from the current index sequencing cycle based on
(i) intensity values of index images from one or more non-current index sequencing cycles, and
(ii) intensity values of index images from the current index sequencing cycle.
11. The artificial intelligence-based method of claim 10, wherein the non-current index sequencing cycles comprise initial index sequencing cycles of the sequencing.
12. The artificial intelligence-based method of claim 10, wherein the non-current index sequencing cycles comprise intermediate index sequencing cycles of the sequencing.
13. The artificial intelligence-based method of claim 10, wherein the non-current index sequencing cycles comprise terminal index sequencing cycles of the sequencing.
14. The artificial intelligence-based method of claim 13, wherein the non-current index sequencing cycles comprise a combination of the initial index sequencing cycles, the intermediate index sequencing cycles, and the terminal index sequencing cycles.
15. The artificial intelligence-based method of claim 10, wherein at least one index image from the non-current index sequencing cycles depicts one or more nucleotides in the detectable signal state.
16. An artificial intelligence-based method of base calling analytes at index sequencing cycles of a sequencing run, the method including:
preprocessing index images generated during the index sequencing cycles using a normalization function that produces a normalized version of an index image from a current index sequencing cycle based on
(i) intensity values of index images from one or more preceding index sequencing cycles,
(ii) intensity values of index images from one or more succeeding index sequencing cycles, and
(iii) intensity values of index images from the current index sequencing cycle;
for a particular analyte being base called at the current index sequencing cycle,
extracting index image patches from normalized versions of the index images from the current, preceding, succeeding index sequencing cycles, such that,

each normalized index image patch depicts intensity emissions of the particular analyte, of some adjacent analytes, and of their surrounding background generated as a result of nucleotide incorporation in corresponding index sequences of the particular analyte and the adjacent analytes during the current index sequencing cycle;

convolving the normalized index image patches through a convolutional neural network and generating a convolved representation; and

base calling the particular analyte at the current index sequencing cycle based on the convolved representation.

17. An artificial intelligence-based method of base calling target sequences and index sequences, wherein the target sequences are derived from a plurality of samples and coupled to the index sequences to form target-index sequences, wherein each index sequence is uniquely associated with a respective sample in the plurality of samples, wherein the target-index sequences are pooled for sequencing during a sequencing run, and wherein the target sequences are sequenced during target sequencing cycles of the sequencing run and the index sequences are sequenced during index sequencing cycles of the sequencing run, the method including:

accessing target images generated for the target sequences during the target sequencing cycles, wherein the target images depict intensity emissions generated as a result of nucleotide incorporation in the target sequences;

preprocessing the target images using a first normalization function that produces a normalized version of a target image from a current target sequencing cycle based only on intensity values of the target image;

processing normalized versions of the target images through a neural network-based base caller and generating a base call for each of the target sequencing cycles, thereby producing target reads for the target sequences;

accessing index images generated for the index sequences during the index sequencing cycles, wherein the index images depict intensity emissions generated as a result of nucleotide incorporation in the index sequences;

preprocessing the index images using a second normalization function that produces a normalized version of an index image from a current index sequencing cycle based on

(i) intensity values of index images from one or more preceding index sequencing cycles,

(ii) intensity values of index images from one or more succeeding index sequencing cycles, and

(iii) intensity values of index images from the current index sequencing cycle;
processing normalized versions of the index images through the neural network-based base caller and generating a base call for each of the index sequencing cycles, thereby producing index reads for the index sequences; and
classifying each target read of a target sequence as belonging to a particular sample in the plurality of samples based on a corresponding index read of an index sequence that is coupled to the target sequence.

18. The artificial intelligence-based method of claim 17, wherein the first normalization function calculates

a lower percentile of the intensity values of the target image, and
an upper percentile of the intensity values of the target image, such that,
in the normalized version of the target image,
a first percentage of normalized intensity values are below the lower percentile,
a second percentage of the normalized intensity values are above the upper percentile, and
a third percentage of the normalized intensity values are between the lower and upper percentiles.

19. An artificial intelligence-based method of base calling target sequences and index sequences, wherein the target sequences are derived from a plurality of samples and coupled to the index sequences to form target-index sequences, wherein each index sequence is uniquely associated with a respective sample in the plurality of samples, wherein the target-index sequences are pooled for sequencing during a sequencing run, and wherein the target sequences are sequenced during target sequencing cycles of the sequencing run and the index sequences are sequenced during index sequencing cycles of the sequencing run, the method including:

accessing target images generated for the target sequences during the target sequencing cycles, wherein the target images depict intensity emissions generated as a result of nucleotide incorporation in the target sequences;

preprocessing the target images using a normalization function that produces a normalized version of a target image from a current target sequencing cycle based on (i) intensity values of target images from one or more preceding target sequencing cycles, (ii) intensity values of target images from one or more succeeding target sequencing cycles, and (iii) intensity values of target images from the current target sequencing cycle;

accessing index images generated for the index sequences during the index sequencing cycles, wherein the index images depict intensity emissions generated as a result of nucleotide incorporation in the index sequences;

preprocessing the index images using the normalization function that produces a normalized version of an index image from a current index sequencing cycle based on (i) intensity values of index images from one or more preceding index sequencing cycles, (ii) intensity values of index images from one or more succeeding index sequencing cycles, and (iii) intensity values of index images from the current index sequencing cycle;

processing normalized versions of the target images through a neural network-based base caller and generating a base call for each of the target sequencing cycles, thereby producing target reads for the target sequences;

processing normalized versions of the index images through the neural network-based base caller and generating a base call for each of the index sequencing cycles, thereby producing index reads for the index sequences; and

classifying each target read of a target sequence as belonging to a particular sample in the plurality of samples based on a corresponding index read of an index sequence that is coupled to the target sequence.

20. The artificial intelligence-based method of claim 19, wherein the normalization function calculates

a lower percentile of (i) the intensity values of the target images from the one or more preceding target sequencing cycles, (ii) the intensity values of the target images from the one or more succeeding target sequencing cycles, and (iii) the intensity values of the target images from the current target sequencing cycle, and

an upper percentile of (i) the intensity values of the target images from the one or more preceding target sequencing cycles, (ii) the intensity values of the target images from the one or more succeeding target sequencing cycles, and (iii) the intensity values of the target images from the current target sequencing cycle, such that,

in the normalized version of the target image,

a first percentage of normalized intensity values are below the lower percentile,

a second percentage of the normalized intensity values are above the upper percentile, and

a third percentage of the normalized intensity values are between the lower and upper percentiles.