

CLAIMS

1. A computer-implemented method of identifying unreliable clusters to improve accuracy and efficiency of base calling, the method including:
 - accessing per-cycle cluster data for a plurality of clusters and for a first subset of sequencing cycles of a sequencing run;
 - base calling each cluster in the plurality of clusters at each sequencing cycle in the first subset of sequencing cycles, including
 - processing the per-cycle cluster data and generating intermediate representations of the per-cycle cluster data, and
 - processing the intermediate representations through an output layer and producing a per-cluster, per-cycle probability quadruple for each cluster and for each sequencing cycle, wherein a particular per-cluster, per-cycle probability quadruple identifies probabilities of a base incorporated in a particular cluster at a particular sequencing cycle being A, C, T, and G;
 - determining a filter value for each per-cluster, per-cycle probability quadruple based on the probabilities it identifies, thereby generating a sequence of filter values for each cluster;
 - identifying those clusters in the plurality of clusters as unreliable clusters whose sequences of filter values contain at least “N” number of filter values below a threshold “M”; and
 - bypassing base calling the unreliable clusters at a remainder of sequencing cycles of the sequencing run, thereby base calling, at the remainder of sequencing cycles, only those clusters in the plurality of clusters that are not identified as the unreliable clusters.
2. The computer-implemented method of claim 1, wherein the filter value for a per-cluster, per-cycle probability quadruple is determined based on an arithmetic operation involving one or more of the probabilities.
3. The computer-implemented method of claims 1-2, wherein the arithmetic operation is subtraction.
4. The computer-implemented method of claims 1-3, wherein the filter value for the per-cluster, per-cycle probability quadruple is determined by subtracting a second highest one of the probabilities from a highest one of the probabilities.
5. The computer-implemented method of claims 1-4, wherein the arithmetic operation is division.

6. The computer-implemented method of claims 1-5, wherein the filter value for the per-cluster, per-cycle probability quadruple is determined as a ratio of the highest one of the probabilities to the second highest one of the probabilities.
7. The computer-implemented method of claims 1-6, wherein the arithmetic operation is addition.
8. The computer-implemented method of claims 1-7, wherein the arithmetic operation is multiplication.
9. The computer-implemented method of claims 1-8, wherein the “N” ranges from 1 to 5.
10. The computer-implemented method of claims 1-9, wherein the “M” ranges from 0.5 to 0.99.
11. The computer-implemented method of claims 1-10, wherein the first subset includes 1 to 25 sequencing cycles of the sequencing run.
12. The computer-implemented method of claims 1-11, wherein the first subset includes 1 to 50 sequencing cycles of the sequencing run.
13. The computer-implemented method of claims 1-12, wherein the output layer is a softmax layer and the probabilities in the per-cluster, per-cycle probability quadruple are exponentially normalized classification scores that sum to unity.
14. The computer-implemented method of claims 1-13, wherein the unreliable clusters are indicative of empty, polyclonal, and dim wells on a patterned flow cell.
15. The computer-implemented method of claims 1-14, wherein the filter values are generated by a filtering function.
16. The computer-implemented method of claims 1-15, wherein the filtering function is a chastity filter that defines chastity as a ratio of a brightest base intensity divided by a sum of the brightest base intensity and a second brightest base intensity.
17. The computer-implemented method of claims 1-16, wherein the filtering function is at least one of a maximum log probability function, a minimum squared error function, average signal-to-noise ratio (SNR), and a minimum absolute error function.

18. The computer-implemented method of claims 1-17, further including:
- determining the average SNR over sequencing cycles in the first subset of sequencing cycles for each cluster based on intensity data in the per-cycle cluster data, wherein the intensity data depicts intensity emissions of clusters in the plurality of clusters and of surrounding background; and
 - identifying those clusters in the plurality of clusters as the unreliable clusters whose average SNR is below a threshold.
19. The computer-implemented method of claims 1-18, further including:
- determining an average probability score for each cluster based on maximum probability scores in per-cluster, per-cycle probability quadruples produced for the sequencing cycles in the first subset of sequencing cycles; and
 - identifying those clusters in the plurality of clusters as the unreliable clusters whose average probability score is below a threshold.
20. A system for improving accuracy and efficiency of neural network-based base calling, the system comprising:
- memory storing, for a plurality of clusters, initial cluster data for initial sequencing cycles of a sequencing run and remainder cluster data for remainder sequencing cycles of the sequencing run;
 - a host processor having access to the memory and configured to execute a detection and filtering logic to identify unreliable clusters;
 - a configurable processor having access to the memory and configured to execute a neural network to produce base call classification scores; and
 - a data flow logic having access to the memory, the host processor, and the configurable processor and configured
 - to provide the initial cluster data to the neural network and cause the neural network to produce initial base call classification scores for the plurality of clusters and for the initial sequencing cycles based on generating initial intermediate representations from the initial cluster data,
 - to provide the initial base call classification scores to the detection and filtering logic and cause the detection and filtering logic to identify unreliable clusters in the plurality of clusters based on generating filter values from the initial base call classification scores,
 - to provide the remainder cluster data to the neural network and cause the neural network to generate remainder intermediate representations from the remainder cluster data, and

to provide data identifying the unreliable clusters to the configurable processor and cause the configurable processor to generate reliable remainder intermediate representations by removing, from the remainder intermediate representations, those portions that result from portions of the remainder cluster data that represent the unreliable clusters.