

CLAIMS

1. A computer-implemented method of generating base calls by a base caller, comprising:

receiving, from a plurality of clusters within a region of a flow cell, a plurality of sensor data representing the plurality of clusters incorporating different nucleotide bases with different labels, wherein the plurality of sensor data is within a first range;

identifying a second range, such that at least a threshold percentage of the plurality of sensor data are within the second range;

mapping at least a subset of the plurality of sensor data, that are within the second range and that represent, from the region of the flow cell, a subset of the plurality of clusters incorporating different nucleotide bases with different labels, to a third range, thereby generating a plurality of normalized sensor data for a target cluster of the plurality of clusters; ~~and~~

processing the plurality of normalized sensor data in a base caller, to ~~call~~ determine, for the target cluster, a ~~corresponding~~-base call and a plurality of quality scores corresponding to the base call; and

remapping each of at least a subset of the plurality of quality scores to a remapped quality score for the base call.

2. The computer-implemented method of claim 1, wherein the second range is fully encompassed within the first range.

3. The [computer-implemented](#) method of claim 1 or 2, wherein one or more outlier sensor data within the first range are absent from the second range of sensor data.

4. The [computer-implemented](#) method of any of claims 1-3, wherein identifying the second range comprises:

identifying, within the first range, a low value, such that a lower threshold percentage of the plurality of sensor data have a value that is lower than the low value; and

identifying, within the first range, a high value, such that an upper threshold percentage of the plurality of sensor data have a value that is higher than the high value,

wherein the second range is defined by the low value and the high value.

5. The [computer-implemented](#) method of claim 4, wherein at least one of the lower threshold percentage or the upper threshold percentage is 0.5% or less.

6. The [computer-implemented](#) method of claim 4, wherein at least one of the lower threshold percentage or the upper threshold percentage is 1.0% or less.

7. The [computer-implemented](#) method of any of claims 4-6, wherein each of the lower threshold percentage and the upper threshold percentage is 0.5% or less.

8. The [computer-implemented](#) method of any of claims 4-6, wherein each of the lower threshold percentage and the upper threshold percentage is 1% or less.

9. The [computer-implemented](#) method of any of claims 4-8, further comprising:

identifying (i) a first outlier sensor data of the plurality of sensor data representing the plurality of clusters incorporating different nucleotide bases with different labels that is lower than the low value and (ii) a second outlier sensor data of the plurality of sensor data representing the plurality of clusters incorporating different nucleotide bases with different labels that is higher than the high value; and

prior to the mapping, assigning the low value to the first outlier sensor data, and assigning the high value to the second outlier sensor data, such that the first outlier sensor data and the second outlier sensor data are within the second range subsequent to the assignment.

10. The [computer-implemented](#) method of any of claims 4-9, further comprising:

identifying (i) a first outlier sensor data of the plurality of sensor data representing the plurality of clusters incorporating different nucleotide bases with different labels that is lower than the low value and (ii) a second outlier sensor data of the plurality of sensor data representing the plurality of clusters incorporating different nucleotide bases with different labels that is higher than the high value; and

excluding the first outlier sensor data and the second outlier sensor data from the subset of the plurality of sensor data representing the subset of the plurality of clusters incorporating different nucleotide bases with different labels during the mapping, for being outside the second range, such that the first outlier sensor data and the second outlier sensor data are not mapped to the third range.

11. The computer-implemented method of any of claims 1-10, wherein mapping at least a subset of the plurality of sensor data representing the subset of the plurality of clusters incorporating different nucleotide bases with different labels comprises:

mapping a first sensor data within the subset from a first value that is within the second range to a second value that is within the third range; and

mapping a second sensor data within the subset from a third value that is within the second range to a fourth value that is within the third range.

~~12. (Original) The method of any of claims 1-11, wherein at least a part of the second range is non-overlapping with the third range.~~

~~12~~¹³. The computer-implemented method of any of claims 1-1~~1~~², wherein individual sensor data of the plurality of sensor data representing the plurality of clusters incorporating different nucleotide bases with different labels comprises corresponding intensity of a corresponding section of an image generated from the flow cell.

~~13~~¹⁴. The computer-implemented method of any of claims 1-1~~2~~³, further comprising:
processing the plurality of normalized sensor data in a base caller, to assign, the corresponding base called for the target cluster, a first quality score indicating a first probability of the corresponding base being an A, a second quality score indicating a second probability of the corresponding base being a C, a third quality score indicating a third probability of the corresponding base being a T, and a fourth quality score indicating a fourth probability of the corresponding base being a G.

~~1415.~~ The computer-implemented method of claim ~~134~~, ~~further comprising~~ ~~wherein: assigning a~~ the plurality of quality scores corresponding to the base call ~~that includes~~ comprise the first quality score, the second quality score, the third quality score, and the fourth quality score; ~~and~~

~~remapping each of at least a subset of the plurality of quality scores to a corresponding remapped quality score.~~

~~1516.~~ The computer-implemented method of claim ~~145~~, further comprising:
quantizing each of a plurality of remapped quality scores to a corresponding one of a plurality of quantized remapped quality score.

16. A non-transitory computer readable storage medium comprising computer program instructions that, when executed on a processor, cause a computing device to:

receive, from a plurality of clusters within a region of a flow cell, a plurality of sensor data representing the plurality of clusters incorporating different nucleotide bases with different labels, wherein the plurality of sensor data is within a first range;

identify a second range, such that at least a threshold percentage of the plurality of sensor data are within the second range;

map at least a subset of the plurality of sensor data, that are within the second range and that represent, from the region of the flow cell, a subset of the plurality of clusters incorporating different nucleotide bases with different labels, to a third range, thereby generating a plurality of normalized sensor data for a target cluster of the plurality of clusters;

process the plurality of normalized sensor data in a base caller, to determine, for the target cluster, a base call and a plurality of quality scores corresponding to the base call; and
remap each of at least a subset of the plurality of quality scores to a remapped quality score for the base call.

17. The non-transitory computer readable storage medium of claim 16, wherein identifying the second range comprises:

identifying, within the first range, a low value, such that a lower threshold percentage of the plurality of sensor data have a value that is lower than the low value; and

identifying, within the first range, a high value, such that an upper threshold percentage of the plurality of sensor data have a value that is higher than the high value, wherein the second range is defined by the low value and the high value.

18. The non-transitory computer readable storage medium of claim 17, further comprising computer program instructions that, when executed on the processor, cause the computing device to quantize each of a plurality of remapped quality scores to a corresponding one of a plurality of quantized remapped quality score.

19. A system comprising:
at least one processor; and
a non-transitory computer-readable medium comprising instructions thereon that, when executed by the at least one processor, cause the system to:

receive, from a plurality of clusters within a region of a flow cell, a plurality of sensor data representing the plurality of clusters incorporating different nucleotide bases with different labels, wherein the plurality of sensor data is within a first range;

identify a second range, such that at least a threshold percentage of the plurality of sensor data are within the second range;

map at least a subset of the plurality of sensor data, that are within the second range and that represent, from the region of the flow cell, a subset of the plurality of clusters incorporating different nucleotide bases with different labels, to a third range, thereby generating a plurality of normalized sensor data for a target cluster of the plurality of clusters;

process the plurality of normalized sensor data in a base caller, to determine, for the target cluster, a base call and a plurality of quality scores corresponding to the base call; and

remap each of at least a subset of the plurality of quality scores to a remapped quality score for the base call.

20. The system of claim 19, further comprising instructions that, when executed by the at least one processor, cause the system to quantize each of a plurality of remapped quality scores to a corresponding one of a plurality of quantized remapped quality score.

~~17. A non-transitory computer readable storage medium impressed with computer program instructions that, when executed on a processor, implement a method comprising:~~

~~receiving, from a plurality of clusters within a region of a flow cell, a plurality of intensity values representing the plurality of clusters incorporating different nucleotide bases with different labels, wherein the plurality of intensity values are within a first range and an individual intensity~~

value depicts a target cluster or an immediate vicinity of the target cluster within the region of the flow cell, the target cluster populated with an unknown nucleotide base;

identifying a second range that includes at least a threshold percentage of the plurality of intensity values representing, from the region of the flow cell, a subset of the plurality of clusters incorporating different nucleotide bases with different labels;

mapping the threshold percentage of the plurality of intensity values that represent the threshold percentage of the plurality of clusters incorporating different nucleotide bases with different labels to a third range that is different from the second range; and

subsequent to the mapping, processing the threshold percentage of the plurality of intensity values representing the subset of the plurality of clusters incorporating different nucleotide bases with different labels, to generate likelihoods of the unknown nucleotide base being an A, C, T, or G.

18. —The non-transitory computer readable storage medium of claim 17, wherein the second range is fully encompassed within the first range.

19. —The non-transitory computer readable storage medium of claim 17 or 18, wherein one or more outlier intensity values within the first range are absent from the threshold percentage of the plurality of intensity values representing the plurality of clusters incorporating different nucleotide bases with different labels.

20. —The non-transitory computer readable storage medium of any of claims 17-19, wherein identifying the second range comprises:

~~identifying, within the first range, a low value, such that a lower threshold percentage of the plurality of intensity values representing the plurality of clusters incorporating different nucleotide bases with different labels have a value that is lower than the low value; and~~

~~identifying, within the first range, a high value, such that an upper threshold percentage of the plurality of intensity values representing the plurality of clusters incorporating different nucleotide bases with different labels have a value that is higher than the high value, wherein the threshold percentage is a sum of the lower threshold percentage and the higher threshold percentage;~~

~~wherein the second range is defined by the low value and the high value.~~

~~21. The non-transitory computer readable storage medium of claim 20, wherein at least one of the lower threshold percentage or the upper threshold percentage is 0.5% or less.~~

~~22. The non-transitory computer readable storage medium of claim 20 or 21, wherein each of the lower threshold percentage and the upper threshold percentage is 1.0% or less.~~

~~23. The non-transitory computer readable storage medium of any of claims 20-22, further comprising:~~

~~identifying (i) a first outlier intensity value of the plurality of intensity values representing the plurality of clusters incorporating different nucleotide bases with different labels that is lower than the low value and (ii) a second outlier intensity value of the plurality of intensity values representing the plurality of clusters incorporating different nucleotide bases with different labels that is higher than the high value; and~~

~~prior to the mapping, assigning the low value to the first outlier intensity value, and assigning the high value to the second outlier intensity value, such that the first outlier intensity value and the second outlier intensity value are within the second range subsequent to the assignment.~~

~~24. — The non-transitory computer readable storage medium of any of claims 20-23, further comprising:~~

~~identifying (i) a first outlier intensity value of the plurality of intensity values representing the plurality of clusters incorporating different nucleotide bases with different labels that is lower than the low value and (ii) a second outlier intensity value of the plurality of intensity values representing the plurality of clusters incorporating different nucleotide bases with different labels that is higher than the high value; and~~

~~excluding the first outlier intensity value and the second outlier intensity value from a subset of the plurality of intensity values representing the plurality of clusters incorporating different nucleotide bases with different labels during the mapping, for being outside the second range, such that the first outlier intensity value and the second outlier intensity value are not mapped to the third range.~~

~~25. — The non-transitory computer readable storage medium of any of claims 17-24, wherein the mapping comprises:~~

~~mapping a first intensity value from a first value that is within the second range to a second value that is within the third range; and~~

~~mapping a second intensity value from a third value that is within the second range to a fourth value that is within the third range.~~

~~26. — The non-transitory computer-readable storage medium of any of claims 17-25, wherein at least a part of the second range is non-overlapping with the third range.~~

~~27. — A system for base calling, comprising:
memory storing images that depict original intensity emissions of a set of clusters, within a region of a flow cell, incorporating different nucleotide bases with different labels, the original intensity emissions generated by clusters incorporating different nucleotide bases with different labels in the set of clusters incorporating different nucleotide bases with different labels during sequencing cycles of a sequencing run;~~

~~a normalization module configured to receive the original intensity emissions and remap the original intensity emissions generated by the clusters incorporating different nucleotide bases with different labels to generate remapped intensity emissions, such that a remapped intensity emission has a different intensity value relative to the original intensity emission; and~~

~~a base caller configured to process the remapped intensity emissions, to generate base calls for the set of clusters, within the region of the flow cell, incorporating different nucleotide bases with different labels.~~