CLAIMS

Claim 1. A computer-implemented method of base calling a target cluster, the method including steps of:

for the target cluster among many clusters is a sequencing cycle that produce differing base-specific intensity distributions,

reading current channel-specific intensities registered for a current sequencing cycle of a sequencing run from a base-specific intensity distribution to which the target cluster is base called at the current sequencing cycle,

reading current channel-specific distribution intensities from a mean intensity centroid of the base-specific intensity distribution,

determining a set of current intensity correction parameters for the current sequencing cycle based on the current channel-specific intensities and the current channel-specific distribution intensities,

determining a set of current accumulated intensity correction parameters for the current sequencing cycle by accumulating the set of current intensity correction parameters and a set of preceding accumulated intensity correction parameters for a preceding sequencing cycle of the sequencing run,

determining a current amplification coefficient that scales a cluster-specific intensity and current channel-specific offset coefficients that offset the cluster-specific intensity for the current sequencing cycle based on the set of current accumulated intensity correction parameters, and

using the current amplification coefficient and the current channel-specific offset coefficients to correct next channel-specific intensities registered for a next sequencing cycle of the sequencing run and generate corrected next channel-specific intensities for the next sequencing cycle; and

base calling the target cluster at the next sequencing cycle based on the corrected next channel-specific intensities.

Claim 2. The computer-implemented method of claim 1, wherein the set of current intensity correction parameters includes the current channel-specific distribution intensities, current channel-specific intensity errors, current distribution centroid-to-origin distance, and current distribution intensity-to-intensity error similarity measure.

Claim 3. The computer-implemented method of any of claims 2, wherein the current channel-specific intensity errors are channel-wise differences between the current channel-specific intensities and the current channel-specific distribution intensities.

Claim 4. The computer-implemented method of any of claims 2-3, wherein the current distribution centroid-to-origin distance is Euclidean distance between the centroid and an origin of a multi-dimensional space that contains the base-specific intensity distribution.

Claim 5. The computer-implemented method of any of claims 2-4, wherein the current distribution intensity-to-intensity error similarity measure is a summation of channel-wise dot products between the current channel-specific distribution intensities and the current channel-specific intensity errors.

Claim 6. The computer-implemented method of any of claims 1-5, wherein the set of current accumulated intensity correction parameters are intensity correction parameter-wise sums of current intensity correction parameters in the set of current intensity correction parameters and preceding accumulated intensity correction parameters in the set of preceding accumulated intensity correction parameters.

Claim 7. The computer-implemented method of claim 1-6, wherein the current channel-specific offset coefficients are configured to be same.

Claim 8. A system comprising:

one or more processors coupled to memory; and

the memory loaded with computer instructions that, when executed by the one or more processors, cause the system to, for a target cluster among many clusters in a sequencing cycle that produce differing base-specific intensity distributions:

read current channel-specific intensities registered for a current sequencing cycle of a sequencing run from a base-specific intensity distribution to which the target cluster is base called at the current sequencing cycle;

read current channel-specific distribution intensities from a centroid of the base-specific intensity distribution;

determine a set of current intensity correction parameters for the current sequencing cycle based on the current channel-specific intensities and the current channel-specific distribution intensities;

determine a set of current accumulated intensity correction parameters for the current sequencing cycle by accumulating the set of current intensity correction parameters and a set of preceding accumulated intensity correction parameters for a preceding sequencing cycle of the sequencing run;

determine a current amplification coefficient that scales a cluster-specific intensity and current channel-specific offset coefficients that offset the cluster-specific intensity for the current sequencing cycle based on the set of current accumulated intensity correction parameters;

use the current amplification coefficient and the current channel-specific offset coefficients to correct next channel-specific intensities registered for a next sequencing cycle of the sequencing run and generate corrected next channel-specific intensities for the next sequencing cycle; and

base call the target cluster at the next sequencing cycle based on the corrected next channel-specific intensities.

Claim 9. The system of claim 8, wherein the set of current intensity correction parameters includes the current channel-specific distribution intensities, current channel-specific intensity errors, current distribution centroid-to-origin distance, and current distribution intensity-to-intensity error similarity measure.

Claim 10. The system of claim 9, wherein current accumulated intensity correction parameters in the set of current accumulated intensity correction parameters include a first common current accumulated intensity correction parameter for the current channel-specific distribution intensities, and a second common current accumulated intensity correction parameter for the current channel-specific intensity errors.

Claim 11. The system of any of claims 9-10, wherein the current channel-specific offset coefficients are channel-wise subtracted from the next channel-specific intensities to generate next channel-specific shifted intensities, and the next channel-specific shifted intensities are divided by the current amplification coefficient to generate the corrected next channel-specific intensities.

Claim 12. The system of any of claims 9-11, further including computer instructions that, when executed by the one or more processors, cause the system to use a weighting function to combine an initial amplification coefficient with the current amplification coefficient, and initial channel-specific offset coefficients with the current channel-specific offset coefficients to generate a weighted current amplification coefficient and weighted current channel-specific offset coefficients for the current sequencing cycle.

Claim 13. The system of any of claims 12, wherein the weighting function applies a minimum weight (wmin) to initial amplification coefficient and the initial channel-specific offset coefficients, and a maximum weight (wmax) to the current amplification coefficient and the current channel-specific offset coefficients, wherein wmin=(1-wmax).

Claim 14. The system of any of claims 13, wherein the maximum weight (wmax) is defined as (c-p)/c, where c is an index for the current sequencing cycle and p is a number between 2 to 7.

Claim 15. The system of any of claims 12-14, further including computer instructions that, when executed by the one or more processors, cause the system to use the weighted current amplification coefficient and the weighted current channel-specific offset coefficients to correct the next channel-specific intensities and generate the corrected next channel-specific intensities.

Claim 16. A non-transitory computer readable storage medium storing instructions that, when executed by at least one processor, cause a system to, for a target cluster among many clusters in a sequencing cycle that produce differing base-specific intensity distributions:

read current channel-specific intensities registered for a current sequencing cycle of a sequencing run from a base-specific intensity distribution to which the target cluster is base called at the current sequencing cycle;

read current channel-specific distribution intensities from a centroid of the base-specific intensity distribution;

determine a set of current intensity correction parameters for the current sequencing cycle based on the current channel-specific intensities and the current channel-specific distribution intensities;

determine a set of current accumulated intensity correction parameters for the current sequencing cycle by accumulating the set of current intensity correction parameters and a set of preceding accumulated intensity correction parameters for a preceding sequencing cycle of the sequencing run;

determine a current amplification coefficient that scales a cluster-specific intensity and current channel-specific offset coefficients that offset the cluster-specific intensity for the current sequencing cycle based on the set of current accumulated intensity correction parameters;

use the current amplification coefficient and the current channel-specific offset coefficients to correct next channel-specific intensities registered for a next sequencing cycle of the sequencing run and generate corrected next channel-specific intensities for the next sequencing cycle; and

base call the target cluster at the next sequencing cycle based on the corrected next channel-specific intensities.

Claim 17. The non-transitory computer readable storage medium of claim 16, further storing instructions that, when executed by at least one processor, cause the system to:

generate, for the current sequencing cycle, a maximum likelihood solution that includes respective current maximum likelihood weights for the current amplification coefficient and the current channel-specific offset coefficients from probability distributions of historical values observed for amplification coefficients and channel-specific offset coefficients in previous sequencing runs;

respectively apply the current maximum likelihood weights to the current amplification coefficient and the current channel-specific offset coefficients to generate a maximum likelihood weighted current amplification coefficient and maximum likelihood weighted current channel-specific offset coefficients for the current sequencing cycle; and

use the maximum likelihood weighted current amplification coefficient and the maximum likelihood weighted current channel-specific offset coefficients to correct the next channel-specific intensities and generate the corrected next channel-specific intensities.

Claim 18. The non-transitory computer readable storage medium of any of claims 16-17, further storing instructions that, when executed by at least one processor, cause the system to:

apply a decay factor to the current intensity correction parameters to generate decayed current intensity correction parameters for the current sequencing cycle; and

determine the current accumulated intensity correction parameters by intensity correction parameter-wise accumulating the decayed current intensity correction parameters and the preceding accumulated intensity correction parameters.

Claim 19. The non-transitory computer readable storage medium of claim 18, wherein the decay factor is kept fixed for a certain number of sequencing cycles of the sequencing run, and exponentially decayed thereafter at each successive sequencing cycle using a decay logic.

Claim 20. The non-transitory computer readable storage medium of claim 19, wherein the decay logic is 1-1/tau, where tau is a predefined number.