MACHINE LEARNING SYSTEM FOR GENOTYPING PCR ASSAYS

BACKGROUND

[0001] Some conventional PCR assay genotyping methodologies (e.g., Taqman®) are based on unsupervised centroid Minimum Cluster Separation Sigma (MCSS) algorithms. A MCSS cutoff (5.0 for example) is empirically selected to tag assays as failure or pass during quality control (QC). However, the hard cutoff means that assays are not classified with nuance. For example, if the cutoff is 5.0, MCSS=5.0 results in a QC pass classification, while MCSS=4.9 results in a QC failure classification. This leads to QC failure of many products that might be acceptable, and thus increases manufacturing loss.

SUMMARY

[0002] A new classification methodology for assay arrays is disclosed based on Support Vector Machine classification and learning, and which may be implemented to genotype cell lines and biological samples. The new methodology can improve the problematic ambiguity of prior QC methods by factoring in historical genotyping results through model training to classify genotypes and to tag qPCR reactions and samples with genotype classifications.

[0002a] One aspect of the present disclosure provides a quality control system including a qPCR system comprising an assay, a storage system coupled to receive first signals, the first signals comprising raw image data resulting from operation of the qPCR system on the assay, and a computing system comprising logic to receive the first signals, receive second signals from the storage system, the second signals comprising labeled data sets labeled by a human classifier, and operate a Support Vector Machine (SVM) to generate classifications for the first signals based on the second signals and to apply the classifications as operational feedback to the qPCR system, wherein the classifications are applied as feedback to adapt the assay or use of the assay in the qPCR system.

[0002b] Another aspect of the present disclosure provides quality control method including operating a qPCR system on an assay to generate first signals, the first signals comprising raw image data resulting from the operation of the qPCR system on the assay, receiving second signals from a storage system, the second signals comprising labeled data sets labeled by a human classifier, operating a Support Vector Machine (SVM) to generate classifications for the first signals based on

the second signals, wherein the SVM is adapted with a kernel comprising $k\left(\overrightarrow{x_i},\overrightarrow{x_j}\right)=exp\left(-\gamma\|\overrightarrow{x_i}-\overrightarrow{x_j}\|^2\right)$, where x is the data vector and γ is a tunable parameter, and a soft margin parameter comprising $min_{w,b}\frac{1}{2}\|w\|_2^2+C\sum_n\zeta_n s.t.\, y_n\left(w^Tx_n+b\right)\geq 1-\zeta_n$, where w and b are parameters for a hyperplane, x_n is the data vector, y_n is the ith target, ζ is a slack variable, and C is a tunable parameter, and applying the classifications to adapt one or both of a process to generate the assay or operate the qPCR system, wherein the classifications are applied as feedback to adapt the manufacture of the assay or use of the assay in the qPCR system.

BRIEF DESCRIPTION OF THE DRAWINGS

[0003] To easily identify the discussion of any particular element or act, the most significant digit or digits in a reference number refer to the figure number in which that element is first introduced.

[0004] FIG. 1 illustrates a process 100 in accordance with one embodiment.

[0005] FIG. 2 illustrates a qPCR system 200 in accordance with one embodiment.

[0006] FIG. 3 illustrates a plate preparation 300 in accordance with one embodiment.

[0007] FIG. 4 illustrates a genotyping system 400 in accordance with one embodiment.

[0008] FIG. 5 illustrates a radial algorithm 500 in accordance with one embodiment.

[0009] FIG. 6 illustrates an SVM qPCR assay model 600 in accordance with one embodiment.

CLAIMS

1. A quality control system comprising:

a qPCR system comprising an assay;

a storage system coupled to receive first signals, the first signals comprising raw image data resulting from operation of the qPCR system on the assay; and

a computing system comprising logic to:

receive the first signals;

receive second signals from the storage system, the second signal comprising labeled data sets labeled by a human classifier;

operate a Support Vector Machine (SVM) to generate classifications for the first signals based on the second signals and to apply the classifications as operational feedback to the qPCR system, wherein the classifications are applied as feedback to adapt the assay or use of the assay in the qPCR system.

2. The quality control system of claim 1, wherein the SVM comprises a radial basis function kernel.

3. The quality control system of claim 2, wherein the kernel comprises:

$k\left(\overrightarrow{x_i}, \overrightarrow{x_j}\right) = exp\left(-\gamma\|\overrightarrow{x_i} - \overrightarrow{x_j}\|^2\right)$, where x is the data vector and γ is a tunable parameter.

4. The quality control system of claim 3, wherein the SVM further comprises a soft margin parameter of:

$min_{w,b}\frac{1}{2}\|w\|_2^2 + C\sum_n \zeta_n s.t.\, y_n\left(w^T x_n + b\right) \geq 1 - \zeta_n$, where w and b are parameters for a hyperplane, x_n is the data vector, y_n is the ith target, ζ is a slack variable, and C is a tunable parameter.

5. The quality control system of claim 1, wherein the storage system and SVM are provided by a cloud server system.

6. The quality control system of claim 1, wherein the SVM is adapted to generate and adapt a model of the assay.

7. The quality control system of claim 6, wherein the model comprises one of SVM linear, polynomial, and radial classifier kernels.

8. A quality control method comprising:

operating a qPCR system on an assay to generate first signals, the first signals comprising raw image data resulting from the operation of the qPCR system on the assay;

receiving second signals from a storage system, the second signals comprising labeled data sets labeled by a human classifier;

operating a Support Vector Machine (SVM) to generate classifications for the first signals based on the second signals, wherein the SVM is adapted with a kernel comprising

$k\left(\overrightarrow{x_i}, \overrightarrow{x_j}\right) = exp\left(-\gamma\|\overrightarrow{x_i} - \overrightarrow{x_j}\|^2\right)$, where x is the data vector and γ is a tunable parameter,

and a soft margin parameter comprising

$min_{w,b}\frac{1}{2}\|w\|_2^2 + C\sum_n \zeta_n s.t. y_n\left(w^T x_n + b\right) \geq 1 - \zeta_n$, where w and b are parameters for a hyperplane, x_n is the data vector, y_n is the ith target, ζ is a slack variable, and C is a tunable parameter; and

applying the classifications to adapt one or both of a process to generate the assay or operate the qPCR system,

wherein the classifications are applied as feedback to adapt the manufacture of the assay or use of the assay in the qPCR system.

9. The quality control method of claim 8, wherein the storage system and SVM are provided by a cloud server system.

10. The quality control method of claim 8, the SVM adapted to generate and adapt a model of the assay.