biological samples in which only one or two clusters occur, a fit to a model may be more difficult, resulting in incorrect genotype calls to be made for at least some samples. For example, a final call in such data sets may depend on the angle of a control sample. In that regard, if the control sample is contaminated, for example, or in any way falsely identified with an incorrect cluster, erroneous calls will be made for every member of that cluster.

[0004] There is a need in the art for a robust analysis of genotype data, in which the optimization is well-defined, and yields a suitable confidence in a final result of assignment of genotype for samples in data sets, where the data sets may be represented by a finite number of clusters of data points based on the ploidy state of the genome of an organism.

SUMMARY

[0004a] According to a first aspect of the present invention, there is provided a computer-readable medium encoded with instructions, executable by a processor, for determining a genotype for a genomic locus in a biological sample, the instructions comprising instructions for: receiving a data set for a genotyping assay, wherein the data set comprises a data point for each sample in a plurality of biological samples; determining an angle for each data point in the data set, wherein a vertex for each angle contains a defined origin; performing an optimal angle configuration search for the data set; said optimal angle configuration search comprising a best fit of the angles determined for each data point in the data set to an angle configuration in a plurality of angle configurations in a data base; and assigning a final genotype classification to each sample in the plurality of biological samples based on the fit of the data set to a best-fit angle configuration, wherein each angle of the best-fit angle configuration is discretely associated with a finite number of allelic combinations defining a genotype classification.

[0004b] According to a second aspect of the present invention, there is provided a computer implemented method for determining a genotype for a genomic locus in a biological sample, the method comprising: receiving a data set for a genotyping assay, wherein the data set comprises a data point for each sample in a plurality of biological samples; processing the data set on a computer to determine the genotype for each sample in a biological, the processing comprising: determining an angle for each data point in the data set, wherein a vertex for each angle contains a defined origin; performing an optimal angle configuration search for the data set; said optimal angle configuration search comprising a best fit of the angles determined for each data point in the data set to an angle configuration in a plurality of angle configurations in a data base; and assigning a final genotype classification to each sample in the plurality of biological samples based on the fit of the data set to the best-fit angle configuration, wherein each angle of the best-fit angle configuration is discretely associated with a finite number of allelic combinations defining a genotype classification.

[0004c] According to a third aspect of the present invention, there is provided a system comprising: a processor; and a memory in communication with the processor; the memory storing instructions for: receiving by a processor a data set for a genotyping assay, wherein the data set comprises a data point for each sample in a plurality of biological samples; determining an angle for each data point in the data set, wherein a vertex for each angle contains a defined origin; performing an optimal angle configuration search for the data set; said optimal angle configuration search comprising a best fit of the angles determined for each data point in the data set to an angle configuration in a plurality of angle configurations in a data base; and assigning a final genotype classification to each sample in the plurality of biological samples based on the fit of the data set to a best-fit angle configuration, wherein each angle of the best-fit angle configuration is discretely associated with a finite number of possible allelic combinations defining a genotype classification.

[0004d] According to a fourth aspect of the present invention, there is provided a computer system comprising: a data base having: records containing information identifying a plurality of angle configurations of signal data for genotyping assays, wherein the angle configurations are generated to cover a defined angle configuration space; and a user interface allowing a user to selectively access the information contained in the records.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] FIG. 1 is a flow chart that depicts various embodiments of methods for the analysis of genotyping data.

[0006] FIG. 2 is a flow chart that depicts various embodiments of methods for the pre-data analysis processing of genotyping data.

[0007] FIG. 3 is an exemplary representation of the assignment of an angle to clusters of data, according to various embodiments of methods for the analysis of genotyping data.

[0008] FIG. 4 is a block diagram of a PCR instrument that may be utilized to process samples analyzed for genotype.

[0009] FIG. 5 is a block diagram of a PCR instrument that may be utilized to process samples analyzed for genotype.

[00010] FIG. 6 is a block diagram that illustrates components of an exemplary computer system that may be utilized in the control and interface of PCR instrumentation.

[00011] FIGS. 7A-7C are graphical representations of angle configurations that may be a subset of angle configurations according to various embodiments of a data base of angle configuration that are possible for genotyping data.

WHAT IS CLAIMED IS:

1. A computer-readable medium encoded with instructions, executable by a processor, for determining a genotype for a genomic locus in a biological sample, the instructions comprising instructions for:

receiving a data set for a genotyping assay, wherein the data set comprises a data point for each sample in a plurality of biological samples;

determining an angle for each data point in the data set, wherein a vertex for each angle contains a defined origin;

performing an optimal angle configuration search for the data set; said optimal angle configuration search comprising a best fit of the angles determined for each data point in the data set to an angle configuration in a plurality of angle configurations in a data base; and

assigning a final genotype classification to each sample in the plurality of biological samples based on the fit of the data set to a best-fit angle configuration, wherein each angle of the best-fit angle configuration is discretely associated with a finite number of allelic combinations defining a genotype classification.

2. The computer-readable medium of Claim 1, wherein the optimal angle configuration search comprising a best fit of the plurality of angles is expressed as a term in a fit score used to evaluate the best fit for the data set.

3. The computer-readable medium of Claim 2, wherein the fit score used to evaluate the best fit for the data set further comprises a term for an assigned probability of occurrence for each angle configuration in the data base of angle configurations.

4. The computer-readable-medium of Claim 3, wherein the term uses a logarithmic expression of the assigned probability.

5. The computer-readable medium of Claim 2, wherein the fit score used to evaluate the best fit for the data set further comprises a term for a Hardy-Weinberg allele frequency.

~~6. The computer-readable medium of Claim 1, wherein an identified baseline is a line containing the defined origin.~~

67. The computer-readable medium of Claim 1, where the origin is defined by a non-template control sample.

78. The computer-readable medium of Claim 1, wherein each angle configuration in the data base comprises three angles.

89. The computer-readable medium of Claim 1, wherein the definable number of allelic combinations is three for a diploid organism.

910. The computer-readable medium of Claim 1, wherein the data base of angle configurations is empirically derived from a data base of genotyping data.

1011. The computer-readable medium of Claim 910, wherein the data base of genotyping data is selected for inherent attributes relating to how genotyping data is generated.

1112. The computer-readable medium of Claim 1, further comprising a step of preprocessing the data.

1213. The computer-readable medium of Claim 1112, wherein preprocessing the data comprises identifying outliers.

1314. The computer-readable medium of Claim 1112, wherein preprocessing the data comprises identifying negative sample controls.

1415. The computer-readable medium of Claim 1112, wherein preprocessing the data comprises identifying non-amplified samples.

1516. The computer-readable medium of Claim 1, further comprising a step of assigning a quality value to the final genotype classification of each sample.

1617. A computer implemented method for determining a genotype for a genomic locus in a biological sample, the method comprising:

receiving a data set for a genotyping assay, wherein the data set comprises a data point for each sample in a plurality of biological samples;

processing the data set on a computer to determine the genotype for each sample in a biological, the processing comprising:

determining an angle for each data point in the data set, wherein a vertex for each angle contains a defined origin;

performing an optimal angle configuration search for the data set; said optimal angle configuration search comprising a best fit of the angles determined for each data point in the data set to an angle configuration in a plurality of angle configurations in a data base; and

assigning a final genotype classification to each sample in the plurality of biological samples based on the fit of the data set to the best-fit angle configuration, wherein each angle of the best-fit angle configuration is discretely associated with a finite number of allelic combinations defining a genotype classification.

17~~18~~. The method of Claim 16~~17~~, further comprising a step of preprocessing the data.

18~~19~~. The method of Claim 16~~17~~, further comprising a step of assigning a quality value to the final genotype classification of each sample.

19~~20~~. The method of Claim 16~~17~~, wherein the optimal angle configuration search comprising a best fit of the plurality of angles is expressed as a term in a fit score used to evaluate the best fit for the data set.

20~~21~~. The method of Claim 19~~20~~, wherein the fit score used to evaluate the best fit for the data set further comprises a term for an assigned probability of occurrence for each angle configuration in the data base of angle configurations.

21~~22~~. The method of Claim 19~~20~~, wherein the fit score used to evaluate the best fit for the data set further comprises a term for a Hardy-Weinberg allele frequency.

22~~23~~. A system comprising:

a processor; and

a memory in communication with the processor; the memory storing instructions for:

receiving by a processor a data set for a genotyping assay, wherein the data set comprises a data point for each sample in a plurality of biological samples;

determining an angle for each data point in the data set, wherein a vertex for each angle contains a defined origin;

performing an optimal angle configuration search for the data set; said optimal angle configuration search comprising a best fit of the angles determined for each data point in the data set to an angle configuration in a plurality of angle configurations in a data base; and

assigning a final genotype classification to each sample in the plurality of biological samples based on the fit of the data set to a best-fit angle configuration, wherein each angle of the best-fit angle configuration is discretely associated with a finite number of possible allelic combinations defining a genotype classification.

2324. A computer system comprising:

a data base having:

records containing information identifying a plurality of angle configurations of signal data for genotyping assays, wherein the angle configurations are generated to cover a defined angle configuration space; and

a user interface allowing a user to selectively access the information contained in the records.

2425. The computer system of Claim 2324, wherein the data base further comprises records containing information identifying for each angle configuration record a probability that the angle configuration may occur.