

Quality value in TaqMan® Genotyper Software



TaqMan® Genotyper Software

TaqMan® Genotyper Software is a SNP genotyping data analysis tool for use with TaqMan® SNP Genotyping Assays (Pre-Designed, Custom, and DME) in combination with 48-, 96- and 384-well microtiter plates, and OpenArray® Genotyping Plates. The software showcases a state-of-the-art genotype calling algorithm, an intuitive user interface, and enhanced multi-plate analysis features to meet the requirements of emerging markets and future research.

TaqMan® Genotyper Software is a SNP genotyping data analysis tool for use with TaqMan® SNP Genotyping Assays. This software is a standalone application to analyze raw data from genotyping experiments performed on an Applied Biosystems® real-time PCR system. The software allows the user to choose between two methods to call genotypes: Autocalling and Classification Scheme. In the Autocalling method, the determining factor influencing the genotype call is a value that the algorithm assigns to each data point: the Quality value. This technical note explains what the Quality value represents and how it is calculated.

What is the Quality value

The Quality value reflects the probability of the genotype call. The Quality value of each data point is listed in the Results table with the corresponding sample ID. In Figure 1, a sample with a higher Quality value of 0.99 will receive a genotype call whereas a sample with a lower Quality value of 0.94, will receive an undetermined call.

#	Sample ID	Call	Manual	Quality	VIC®	FAM™	ROX™	Task	Ger
1	NA17102	Undeter...		0.94	3,168.232	5,365.722	2,011.612	Unknown	Unkn
2	NA17103	FAM/FAM		0.99	1,319.276	10,544....	2,225.646	Unknown	Unkn
3	NA17104	FAM/FAM		0.99	1,928.260	12,053....	2,364.296	Unknown	Unkn
4	NA17105	FAM/FAM		0.99	1,411.921	12,104....	2,205.754	Unknown	Unkn
5	NA17106	FAM/FAM		0.99	1,567.430	12,643....	2,166.849	Unknown	Unkn

Figure 1. The Quality value of each data point is listed in the Results table. For each sample, the Results table displays the genotype call, Quality value, and the VIC®, FAM®, and ROX™ dye values.

When the software analyzes genotyping data, the algorithm applies a default threshold for the Quality value (Figure 2). This default value is specified in the Analysis Settings on the QC Settings tab. A data point with a Quality value lower than the default threshold does not pass quality specifications and receives an undetermined call. Lowering the default threshold may result in an undetermined call being changed to a genotype call.

Details of the Quality value calculation

Each data point is given a Quality value, which reflects the probability of the genotype call. The Quality value takes into account the angle (A) and the amplitude (B) from the NTC (Figure 3). The angle is a measure of the cluster separation, and the amplitude is a measure of the signal intensity.

Each data point can belong to either of the class labels VIC, HET, FAM, INVALID, NOAMP, and OUTLIER in the angle and amplitude space. Each data point receives a score with respect to each of these categories in angle and amplitude space respectively. The amplitude and angle scores are added in quadrature to get the score for the category. For each data point, the following scores are calculated:

$$\text{VIC score} = \sqrt{(\text{VIC amplitude score})^2 + (\text{VIC angle score})^2}$$

$$\text{FAM score} = \sqrt{(\text{FAM amplitude score})^2 + (\text{FAM angle score})^2}$$

$$\text{HET score} = \sqrt{(\text{HET amplitude score})^2 + (\text{HET angle score})^2}$$

$$\text{NOAMP score} = \sqrt{(\text{NOAMP amplitude score})^2 + (\text{NOAMP angle score})^2}$$

$$\text{OUTLIER score} = \sqrt{(\text{OUTLIER amplitude score})^2 + (\text{OUTLIER angle score})^2}$$

The class label that has the maximum score value is assigned to the data point. For example, if the maximum score is the VIC score, the data point is classified as a VIC dye.

The quality value is expressed as a fraction representing how strongly a given data point belongs to a given category label. For the above data point which had its VIC Score as the maximum score, the quality value is calculated as:

$$\text{Quality value} = \frac{\text{VIC Score}}{\text{VIC Score} + \text{FAM Score} + \text{HET Score} + \text{NOAMP Score} + \text{OUTLIER Score}}$$

Some business rules also apply:

1. Designated no-amp points get a confidence of 1.
2. Invalid points get a confidence of 0.

A

Analysis Settings

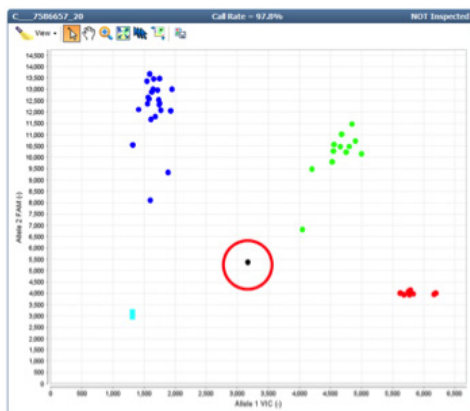
Set the default analysis settings.

Call Settings Control Identifiers QC Settings

Well-Level QC Flags

Use	Icon	Flag Name	Description	Value	Condition	Threshold
<input checked="" type="checkbox"/>		Failed Control	The control call does not match the predefined expect...			
<input checked="" type="checkbox"/>		Genotype Quality Low	The quality value of the data point's genotype determi...	Quality	<	0.95
<input checked="" type="checkbox"/>		Low ROX™ Intensity	The ROX™ dye intensity for the data point is below th...	ROX™	<	800
<input checked="" type="checkbox"/>		NTC FAM™ Intensity High	The FAM™ dye intensity for the no template control is ...	FAM™	>	4,000
<input checked="" type="checkbox"/>		NTC VIC® Intensity High	The VIC® dye intensity for the no template control is a...	VIC®	>	3,000
<input checked="" type="checkbox"/>		Reference Sample Discordance	The sample call does not match the reference sample c...			
<input checked="" type="checkbox"/>		Replicate Sample Discordance	The sample call does not match the replicate sample call.			

B



C

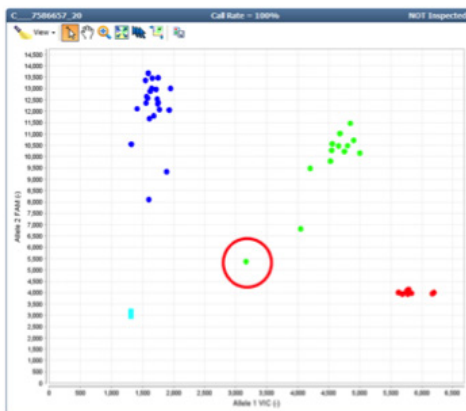


Figure 2. The default threshold for the Quality value is 0.95 (A). This threshold is listed in the Analysis Settings on the QC Settings tab. Lowering the default threshold can increase the sample call rate. The circled sample in the scatter plot (B) is assigned an undetermined call when the default threshold of 0.95 is applied for analysis, but it is assigned a heterozygous call (C) when the Quality value threshold is lowered to 0.80.

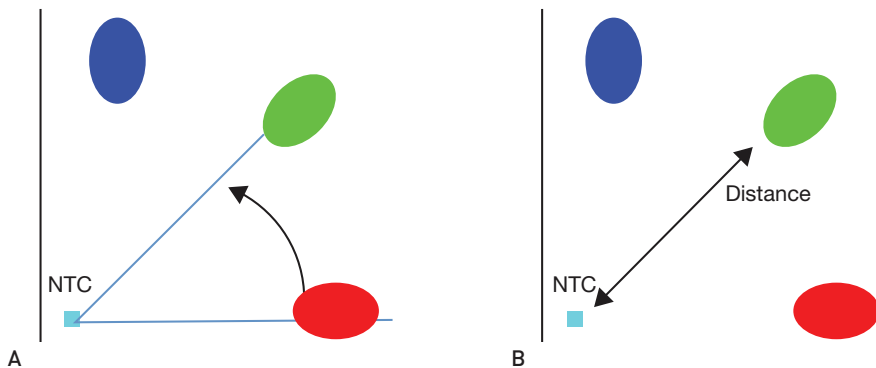


Figure 3. Illustration of angle (A) and amplitude (B) calculated from the NTC. Good cluster separation means good separation of data points from other clusters (angle) and leads to good angle scores. Good signal intensity means good separation of data points from the NTC and generates good amplitude scores.

The amplitude and angle scoring process

Angle (A) and amplitude (B) for each data point are calculated by the formulas below.

$$\text{Angle} = \tan^{-1} \frac{\text{FAM Rn} - \text{mean NTC FAM Rn}}{\text{VIC Rn} - \text{mean NTC VIC Rn}}$$

$$\text{Amplitude} = \sqrt{(\text{FAM Rn} - \text{mean NTC FAM Rn})^2 + (\text{VIC Rn} - \text{mean NTC VIC Rn})^2}$$

The algorithm computes the models for the distribution of angles and amplitudes for each of the class labels. The score is roughly an estimate of how closely a given data point belongs to those underlying models of angle and amplitude distributions for each genotype cluster.

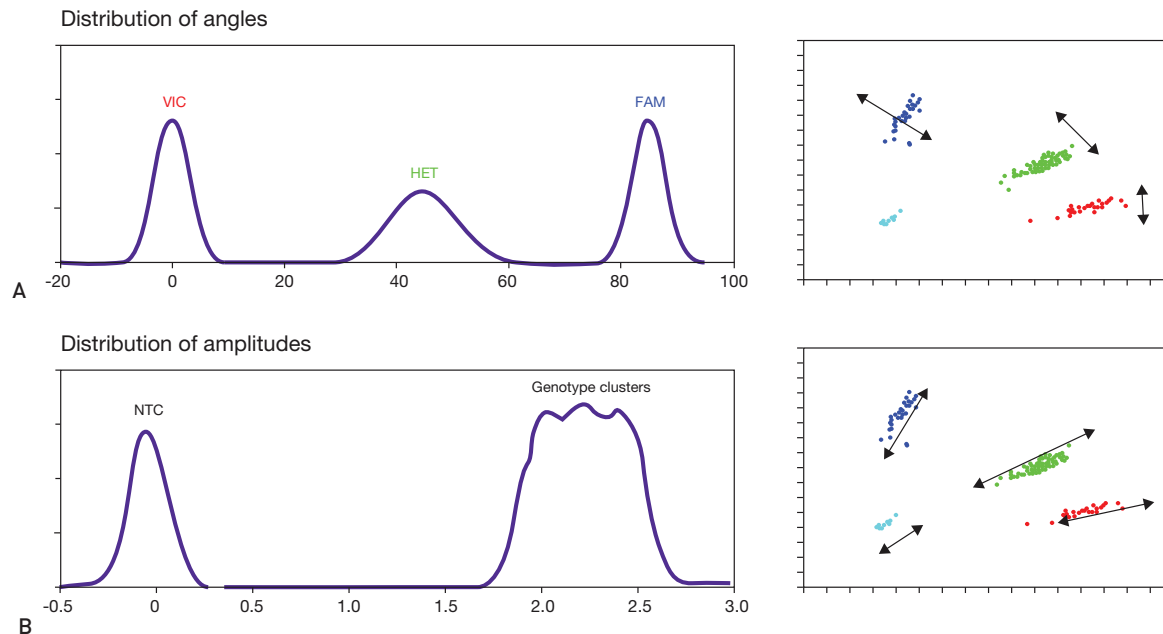


Figure 4. Distribution of angles (A) and amplitudes (B). The distribution of angles (A) shows three well separated clusters. The distribution of amplitudes (B) shows three clusters with good signal intensity that are well separated from the NTC.

The example in Figure 2 shows a highlighted data point with good cluster separation but poor signal intensity. The poor signal intensity translates to a poor amplitude score and lowers the Quality value below the default threshold. In the distribution shown in Figure 4B, this data point would map between the distributions of NTC and genotype clusters. When the default threshold of 0.95 is lowered to 0.80, the sample is assigned a heterozygous call.

Conclusion

This technical note explains how the Quality value in TaqMan® Genotyper Software is calculated. The Quality value is an estimate of how closely a given data point belongs to the underlying model of angle and amplitude distribution for each genotype cluster.

Download TaqMan® Genotyper Software at lifetechnologies.com/genotyper

