User guide

Analysis of end-point genotyping data using cluster plots

For Research Use Only. Not for use in diagnostic procedures.

BIOSEARCH™ TECHNOLOGIES

GENOMIC ANALYSIS BY LGC

User guide

Analysis of end-point genotyping data using cluster plots

Contents

User guide

Analysis of end-point genotyping data using cluster plots

Table of figures

User guide

Analysis of end-point genotyping data using cluster plots

1. Introduction

End-point genotyping assays are based on competitive allele-specific PCR and enable bi-allelic scoring of single nucleotide polymorphisms (SNPs) and insertions and deletions (InDels) at specific loci. LGC, Biosearch Technologies offers three options for end-point genotyping chemistry;

a) KASP™, which comprises of three SNP-specific oligonucleotides that are contained within the KASP Assay Mix, and is used with KASP Master Mix.
b) BHQ™-labelled hydrolysis probe-based SNP genotyping, which utilises two BHQplus™ probes and two primers for each assay, or ValuMix Assays for SNP Genotyping. Both of these may be used with BHQ Probe Master Mix.
c) BHQplex™ CoPrimers™ which are fluorescently-labelled probes, consisting of a 'Primer' sequence and a 'Capture' sequence, connected by a flexible linker to enable cooperative binding to the target of interest. It is recommended to use these with BHQ Probe Master Mix.

Detailed explanations of how each of these three chemistries work can be found in Appendices 1, 2 and 3 respectively. The required combination of oligonucleotides and/or probes required to genotype one individual SNP with each of these chemistries will be referred to as an 'assay' within this document.

Reaction components are combined and added to DNA samples. A thermal cycling reaction is then performed followed by an end-point fluorescent read. Bi-allelic discrimination is achieved through the competitive binding of two allele-specific oligonucleotides (primers and/or probes depending on the chemistry used), each associated with a different fluorescent molecule.

Once the end-point genotyping reactions are complete, and the resulting final fluorescence values have been measured, the raw data must be interpreted to enable genotypes to be assigned to the DNA samples. This is typically achieved by viewing the fluorescence values for each sample on a Cartesian plot. These plots are often referred to as cluster plots or scatter plots.

The purpose of this document is to provide an explanation of how to analyse end-point genotyping data using a cluster plot, and to provide troubleshooting guidance for common issues that can arise.

2. How to visualise genotyping results using a cluster plot

2.1. Data points and axis nomenclature

The fluorescent signal from each individual DNA sample is represented as an independent data point on a cluster plot.

For KASP, one axis is used to plot the FAM™ fluorescence value (typically the X axis) and the second axis is used to plot the HEX™ fluorescence value (typically the Y axis) for each sample.

For BHQplus probes and BHQplex CoPrimers, one axis is used to plot the fluorescence from the X-allele specific probe (typically the X axis) and the second axis is used to plot the fluorescence from the Y-allele specific probe (typically the Y axis) for each sample.

For ease of description and illustration in this troubleshooting guide, all subsequent references and cluster plot diagrams will refer to “X-signal” (for signal generated by FAM when genotyping with KASP and for the X-allele specific probe signal when genotyping with BHQplus probes or BHQplex CoPrimers) and “Y-signal” (for signal generated by HEX when genotyping with KASP and for the Y-allele specific probe signal when genotyping with BHQplus probes or BHQplex CoPrimers).

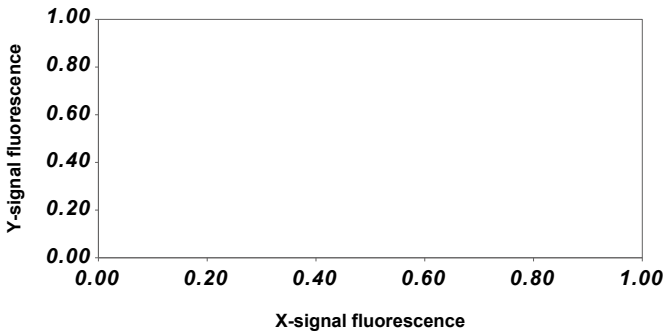


Figure 1. The axes used on a typical genotyping cluster plot.

2.2. Homozygous samples

A sample that is homozygous for the allele reported by the X-signal oligonucleotide will only generate X-signal fluorescence during the end-point genotyping reaction. This data point will be plotted close to the X axis, representing high X-signal and no Y-signal (blue data points in Figure 2).

A sample that is homozygous for the allele reported by the Y-signal oligonucleotide will only generate Y-signal fluorescence during the end-point genotyping reaction. This data point will be plotted close to the Y axis, representing high Y-signal and no X-signal (red data points in Figure 2).

2.3. Heterozygous samples

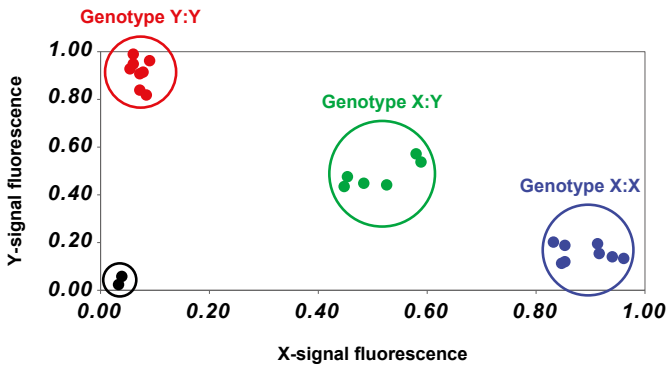
A sample that is heterozygous will contain both the allele reported by the X-signal oligonucleotide and the allele reported by the Y-signal oligonucleotide. This sample will generate half as much X-signal fluorescence and half as much Y-signal fluorescence as the samples that are homozygous for these alleles. This data point will be plotted in the centre of the plot, representing half X-signal and half Y-signal (green data points in Figure 2).

2.4. No template controls

To ensure the reliability of the results, an end-point genotyping reaction without any template DNA must be included as a negative control. This is typically referred to as a no template control (or NTC). The NTC will not generate any fluorescence and the data point will therefore be plotted at the origin of the plot (black data points in Figure 2).

2.5. Interpretation of genotyping clusters

All samples that have the same genotype will have generated similar levels of fluorescence and will therefore all appear close together on the plot. Based on the relative position of these clusters, it is possible to determine the genotype of all the data points. It is important that a sufficient number of individual samples are included in an end-point genotyping reaction plate to ensure that there are enough data points on the Cartesian plot to allow cluster analysis.



Legend	
●	Homozygous for allele X, reported by X-signal oligonucleotide
●	Homozygous for allele Y, reported by Y-signal oligonucleotide
●	Heterozygous, contains allele X and allele Y
●	No template control (NTC)

Figure 2. A typical genotyping cluster plot.

Each data point represents the fluorescence signal of an individual DNA sample. Samples of the same genotype will have generated similar levels of fluorescence and will therefore cluster together on the plot. Blue data points = homozygous for X-allele; red data points = homozygous for Y-allele; green data point = heterozygous; black data points = no template controls (NTCs).

3. Common issues with interpretation of genotyping cluster plots

3.1. Too few individual data points

Biosearch Technologies recommends that a minimum of 22 DNA samples plus 2 NTCs are run per assay per plate to enable cluster analysis to be performed effectively. Running fewer samples than the recommended 22 (plus 2 NTCs) can make it difficult to score genotypes reliably, and may cause misinterpretation of genotyping results.

Figure 3 provides an example of how data can be misinterpreted when too few samples are run; the limited number of data points on Figure 3A affects the scaling of the Y axis, and hence suggests that one of the data points is heterozygous (green triangle). When sufficient data points are included (Figure 3B), the presence of true heterozygote data points (green circles) make it clear that this potential heterozygous sample is actually homozygous for allele X (blue triangle).

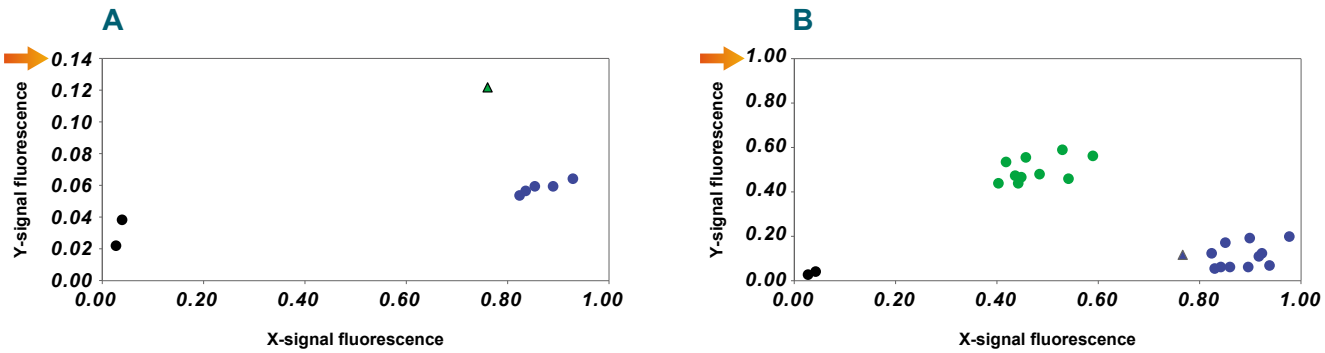


Figure 3. Too few individual data points.

Plot A shows a cluster plot where only six DNA samples have been genotyped. There is separation between the blue circles and the green triangle, and the results suggest that the green triangle is a heterozygote. Plot B shows the same data, but including the results for an additional sixteen DNA samples. The data point represented with a green triangle in A is identified with a blue triangle in B. By including the required number of DNA samples, the genotyping clusters are sufficiently distinct and the triangular data point now clearly belongs to the homozygous cluster (blue). The orange arrows (on A and B) indicate the effect that the number of plotted data points has on the automatic scaling of the Y axis.

3.2. Scaling of X and Y axes is not comparable

Of the two reporting fluorescent molecules used in end-point genotyping chemistry, one may produce a greater signal than the other. This difference in relative fluorescence does not negatively affect the genotyping data when viewed as a cluster plot as the X and Y axes of the plot will be scaled independently to adjust for the different maximum signal produced by each of the fluorophores.

The software supplied with a number of widely available qPCR instruments can sometimes automatically scale the X and Y axes so that they are not comparable. This can affect the view of the genotyping clusters and may cause misinterpretation of genotyping results. To prevent this issue, manually inspect the minimum and maximum values for the X and Y axes. If they differ considerably, rescale the axes so that both have the same minimum and maximum value.

Figure 4 provides an example of how data can be misinterpreted when the X and Y axes are not scaled comparably; on Figure 4A the two clusters present (red and blue) can be interpreted as the two homozygote genotypes, yet once the X axis is scaled correctly to match the Y axis (Figure 4B) it is clear that the blue homozygous cluster (genotype X:X) actually represents heterozygote samples (green, genotype X:Y).

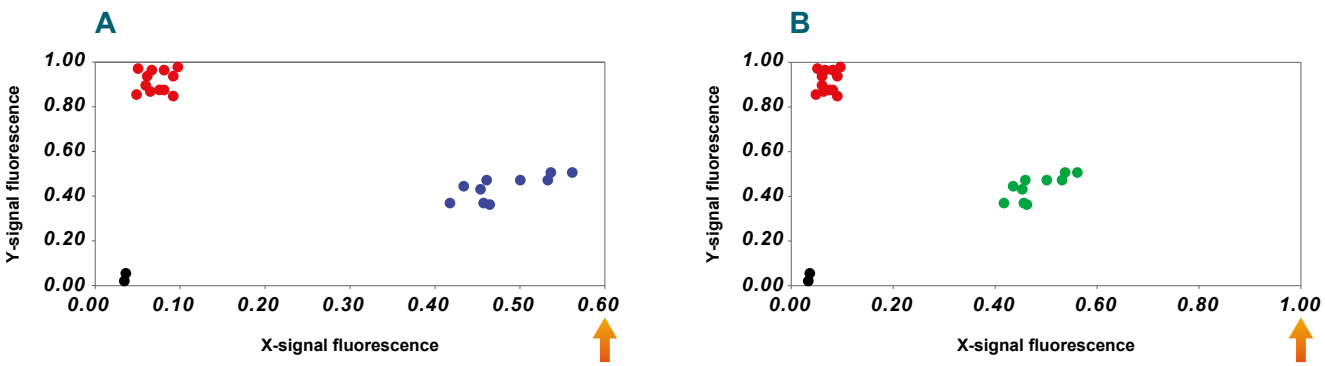


Figure 4. Scaling of X and Y axes is not comparable (polymorphic results).
Plot A shows a cluster plot with two clear genotyping clusters; these have been assigned as homozygous for allele X (blue data points) and homozygous for allele Y (red data points). In plot A, the X and Y axes are not scaled comparably; X has a maximum of 0.60 whilst Y has a maximum of 1.00. Plot B shows the same data as plot A, but with the X axis scaled correctly to match the Y axis. It is then clear that the blue cluster actually represents samples with a heterozygous genotype (green data points).

The scaling of the axes can be a particular issue if all of the data points are the same genotype e.g. monomorphic for the X-signal. In this scenario, only X-allele signal will have been generated from each sample and hence the X-allele axis may be scaled to have a much higher maximum than the Y-allele axis. This can result in monomorphic results appearing as very scattered across the plot (Figure 5A) – rescaling the axes can improve data visualisation and hence the ability to confidently assign a monomorphic genotype (Figure 5B).

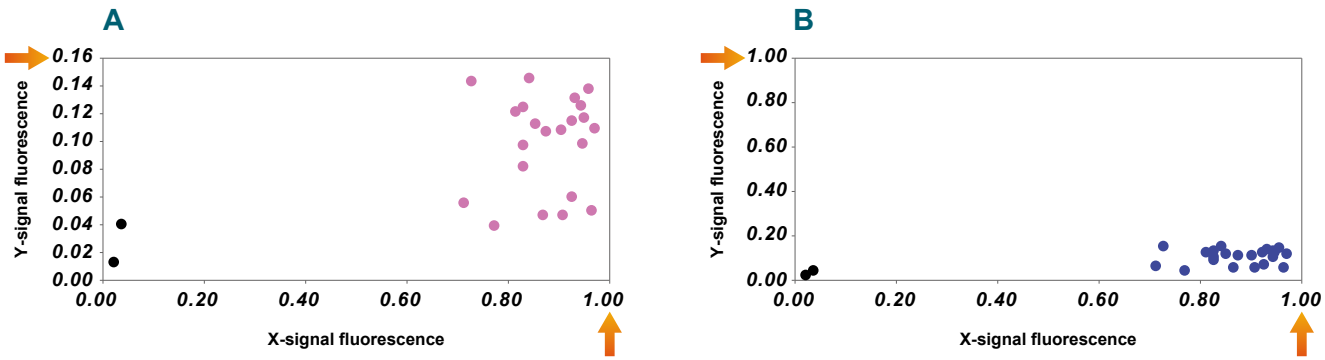


Figure 5. Scaling of X and Y axes is not comparable (monomorphic results).
Plot A shows a cluster plot where the data points do not appear to be forming into clusters (pink, no genotype assigned). The orange arrows on plot A indicate the maximum fluorescence values for the X and Y axes; in this case the X axis has a maximum of 1.00 whilst the Y axis has much smaller maximum of 0.16. Plot B shows the same data as plot A, but with the Y axis re-scaled to have the same maximum fluorescence value as the X axis (1.00, as indicated by the orange arrow). By re-scaling the Y axis to match the X axis, the genotyping data points are positioned correctly and it is then clear that all of the data points belong to one cluster and represent samples with a homozygous genotype (blue cluster).

3.3. Reliance on the software's in-built auto calling option

The software packages supplied with a number of widely available qPCR instruments typically have the ability to automatically assign genotypes to data points based on their position on the cluster plot. Please be aware that the algorithms used by these packages for data analysis will differ in robustness, and no software should be considered as 100% reliable. Reliance on the auto called genotypes with no manual review may result in misinterpretation of genotyping results.

User guide

Analysis of end-point genotyping data using cluster plots

Biosearch Technologies strongly recommends that genotyping data is always visually inspected and that results are sense-checked to ensure that they meet expectations. Ensuring that the X and Y axes on the cluster plot are scaled comparably can aid data interpretation. The inclusion of positive controls (samples of known genotype) on each reaction plate can help to provide confidence in results that are auto called by the software, and will ensure that there are always three clusters present on the cluster plot. Figures 6A and 6B demonstrate the effect of axes scaling, and Figure 6C shows the benefits of including positive controls on each reaction plate.

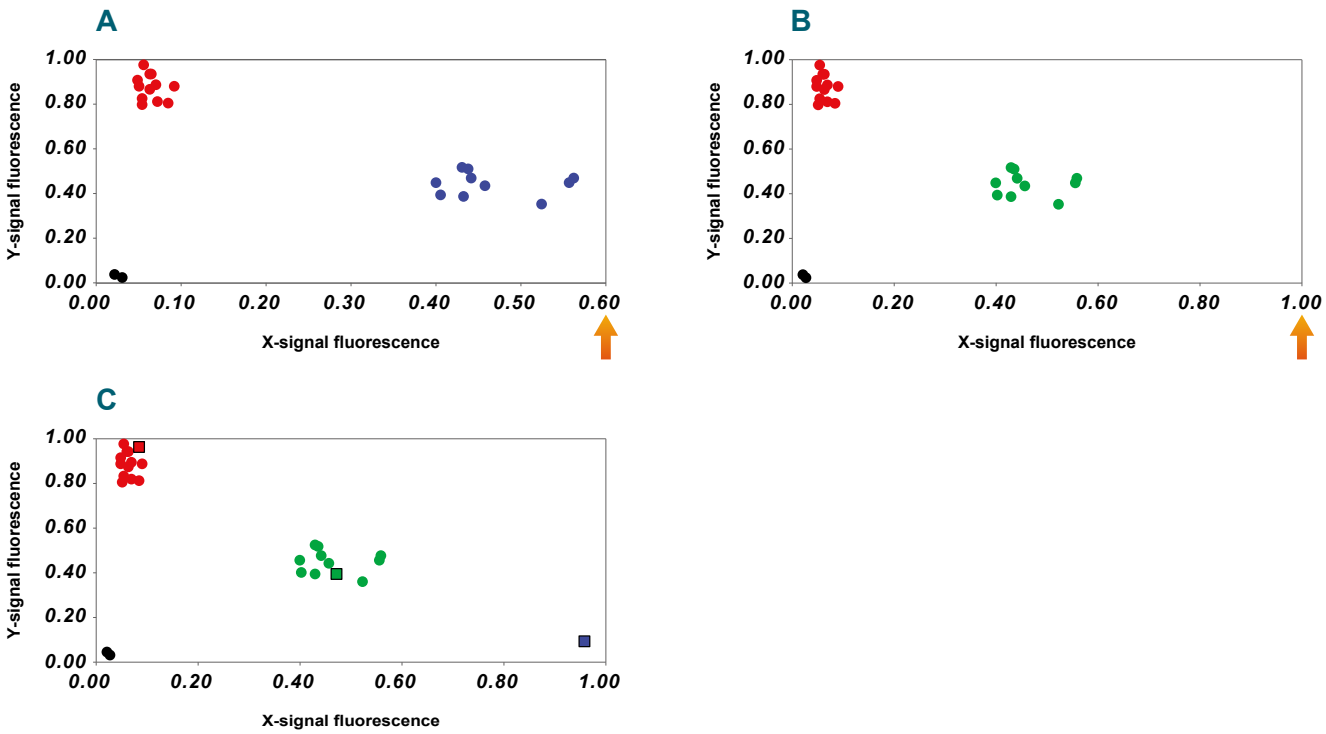


Figure 6. Reliance on the software's in-built auto calling option.

Plot A shows a cluster plot with two clear genotyping clusters that have been auto called using qPCR instrument software. The orange arrow on plot A indicates the maximum fluorescence value for the X axis, in this case 0.60. Plot B shows the same data as plot A, but with the X axis re-scaled to have the same maximum fluorescence value as the X axis (1.00, as indicated by the orange arrow). Rescaling the axes ensures that the genotyping clusters are positioned correctly – in this case, the blue cluster from plot A is now in the centre of the plot and has been manually assigned as a heterozygote cluster (green data points). Plot C shows the same data as plot A and plot B, but with three additional data points, indicated by square markers. These represent a positive control sample for each genotype. Inclusion of these positive controls ensures that the three possible genotypes are represented and can hence improve the accuracy of auto calling features that are in-built in qPCR instrument software.

3.4. Combining data from multiple reaction plates on the same plot

Biosearch Technologies does not recommend combining data from multiple reaction plates. This is because the fluorescence generated, and the resultant scaling of the cluster plots, will differ from plate to plate. Viewing combined data on one plot is likely to make the genotyping clusters less defined, and will affect interpretation of the results.

Figure 7 shows data for the same end-point assay run on two separate reaction plates; viewing each set of data independently (Figures 7A and 7B) gives three tight genotyping clusters. When the data are plotted together on the same plot, however, the clusters become difficult to interpret (Figure 7C).

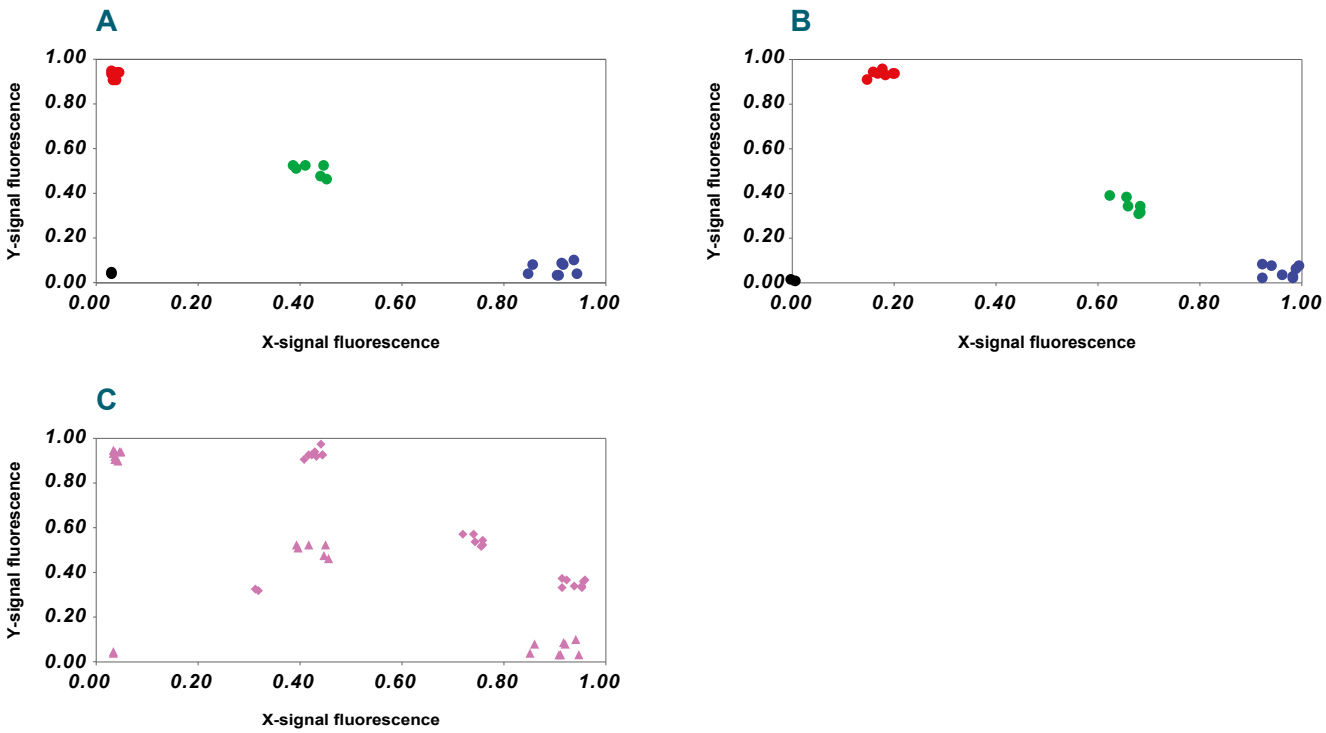


Figure 7. Combining data from multiple reaction plates on the same plot.
Plot A shows the data for an individual end-point assay, generated on reaction plate A. Plot B shows the data for the same end-point assay, generated on reaction plate B. In each case, the cluster plot axes have been scaled appropriately for the signal values generated on each reaction plate and, on both plots, clear clusters are visible and can be assigned genotyping calls. Plot C illustrates the effect of plotting data from two different reaction plates on the same cluster plot; the axes have been scaled based on all the fluorescent signal values generated across the two reaction plates (Triangular data points are derived from Plot A, square data points are derived from Plot B). Variations in the fluorescence generated between plates have resulted in multiple clusters being present on the plot. Consequently, it is no longer possible to confidently assign genotyping calls.

3.5. Combining data from multiple end-point assays on one plot

Biosearch Technologies does not recommend combining data from different end-point assays on the same cluster plot. Each individual end-point assay will amplify template DNA at a slightly different rate, and will subsequently generate differing levels of fluorescent signal. Plotting data from more than one assay on the same plot does not provide any meaningful comparison, and will result in the data being very difficult to interpret (Figure 8).

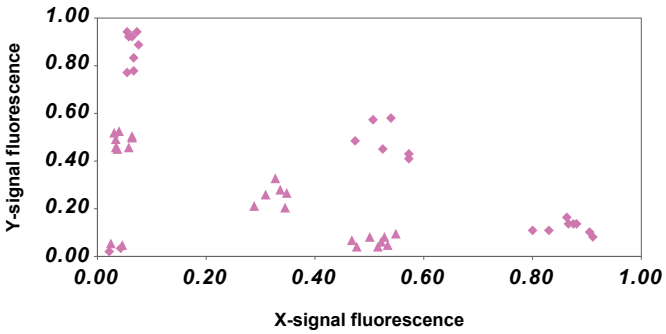


Figure 8. Combining data from two end-point assays on one plot.

This cluster plot contains data points from two different end-point assays. One assay has generated greater fluorescent signal values than the other assay and, as a result, there are multiple clusters present on the plot. Consequently, data interpretation and assigning of genotyping calls is much more difficult.

Figure 9 illustrates the effect of viewing data from multiple assays on the same cluster plot; both data quality and ease of interpretation are drastically affected. In Figure 9A, the pink smear of data points represents data from four different end-point assays and the ability to confidently assign genotyping calls is reduced. Figures 9B–9E identify the data points for each of the individual assays and, by viewing this data individually, the three genotyping clusters are much more apparent. All four assays have generated good genotyping data, with three clear clusters, and if the data points for each assay were viewed on an individual plot then it would be straightforward to confidently assign genotyping calls.

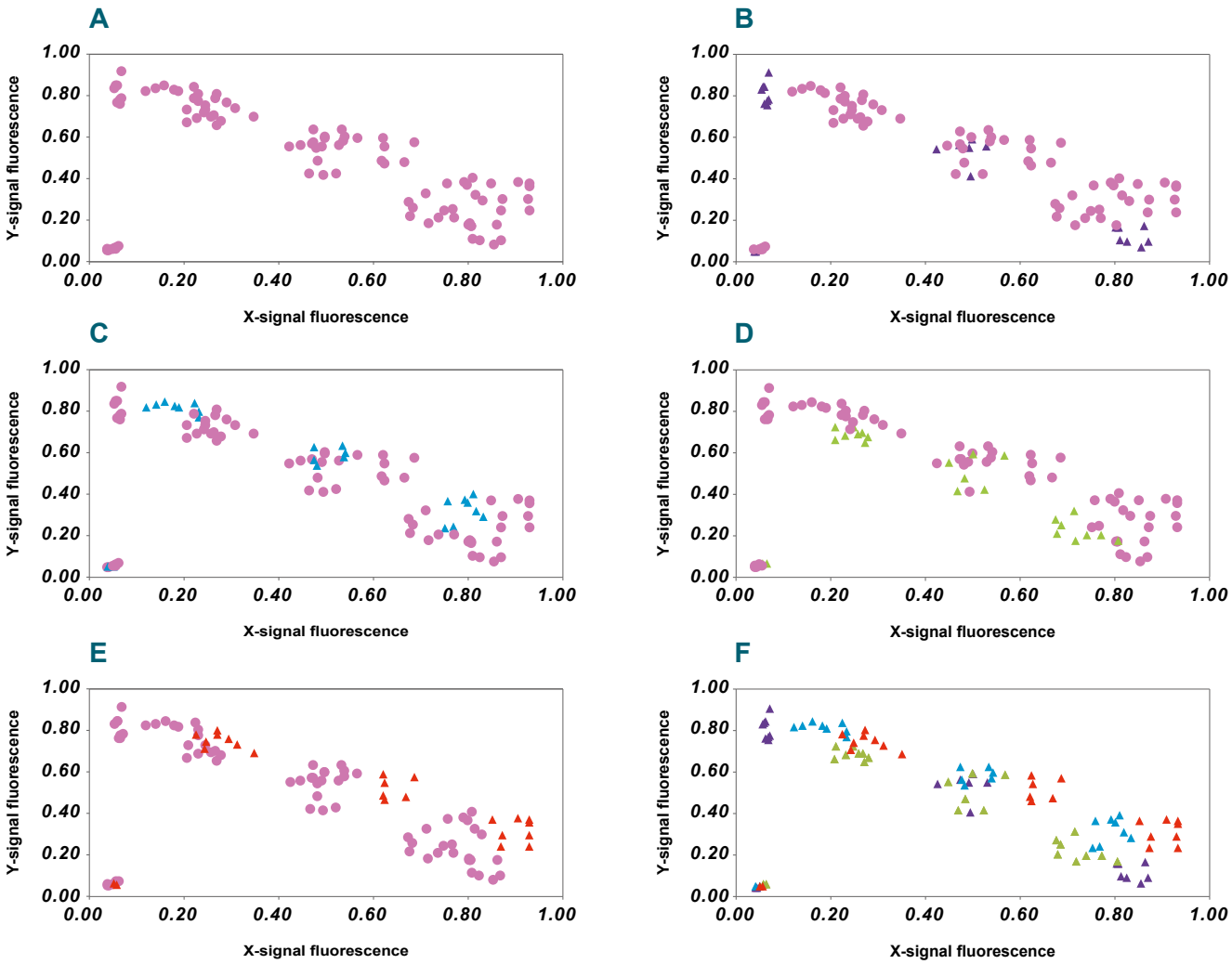


Figure 9. Combining data from multiple end-point assays on one plot.
Plot A shows the data for multiple end-point assays on a single cluster plot. The data points form a smear across the plot and consequently it is not possible to confidently assign genotyping calls. Plots B, C, D and E identify the data for an individual assay using a unique colour for the assay-specific data points (purple, blue, green and red respectively). In each case, three distinct clusters are clearly visible and it would be possible to confidently assign genotyping calls if these data points were viewed on their own cluster plot. Plot F shows the same data as plot A, but with the data points for each end-point assay identified by their unique colour.

4. Summary

This document has outlined the key factors to consider when viewing end-point genotyping data on a cluster plot. When viewing and analysing your data, ensure that you follow these guidelines:

- Include a minimum of 22 data points (plus 2 NTCs) on each plot to facilitate cluster analysis.
- Check the scaling of the X and Y axes and, if required, rescale so that they are comparable. This will ensure that clusters are correctly interpreted.
- Where possible, include positive controls for each genotype as this will facilitate data analysis.
- Do not combine data from different reaction plates on the same plot, even if the results are for the same end-point assay.
- Do not view results for more than one end-point assay on an individual cluster plot – always view results assay by assay.
- Always sense-check genotypes that are automatically assigned by the instrument software as they may need to be manually adjusted.

5. Further support

For any queries about this guide or running end-point genotyping reactions in your laboratory, please contact the technical support team: techsupport@lgcgroup.com.

6. Appendices

6.1. Appendix A: Genotyping with KASP

KASP end-point genotyping chemistry requires a SNP-specific KASP Assay mix that contains three assay-specific non-labelled oligonucleotides; typically two allele-specific forward primers and one common reverse primer. The allele-specific primers each harbour a unique tail sequence that corresponds with a universal FRET (fluorescence resonant energy transfer) cassette; one labelled with FAM dye and the other with HEX dye. The KASP Master Mix contains the universal FRET cassettes, ROX™ passive reference dye, *Taq* polymerase, free nucleotides and $MgCl_2$ in an optimised buffer solution.

During thermal cycling, the relevant allele-specific primer binds to the template and elongates, thus attaching the tail sequence to the newly synthesised strand. The complement of the allele-specific tail sequence is then generated during subsequent rounds of PCR, enabling the FRET cassette to bind to the DNA. The FRET cassette is no longer quenched and emits fluorescence. Bi-allelic discrimination is achieved through the competitive binding of the two allele-specific forward primers.

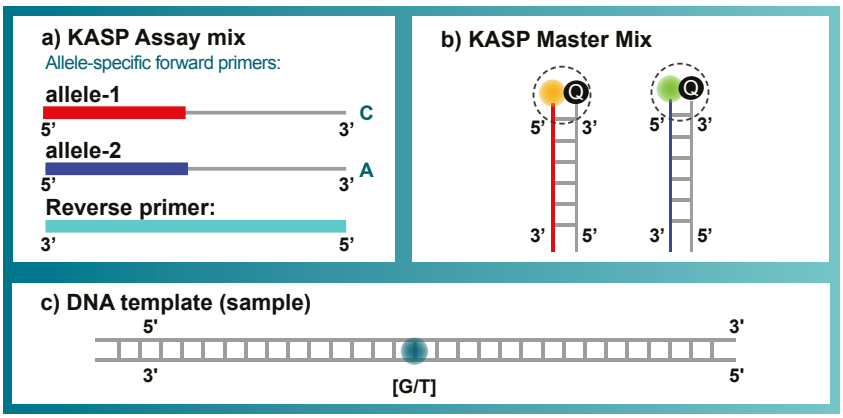
A schematic of how KASP chemistry works is shown in Figure 10.

1: Assay components.

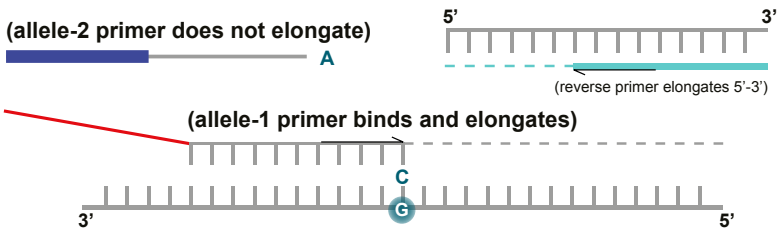
a): KASP Assay mix consists of 2 allele specific primers and 1 reverse primer.

b): KASP Master Mix contains universal fluorescent probes, *Taq* polymerase and dNTP's in an optimised buffer solution.

c): Sample DNA DNA contains the SNP of interest.



2. Denatured template and annealing components – PCR round 1:



In the first round of PCR, one of the allele-specific primers matches the target SNP and with the common reverse primer, amplifies the target region.

Legend

	Allele-1 tail FAM-labelled oligo sequence
	Allele-2 tail HEX-labelled oligo sequence
	Common reverse prime
	FAM dye
	HEX dye
	Target SNP
Q	Quencher

3. Complement of allele-specific tail sequence generated – PCR round 2:

(Reverse primer binds, elongates and makes a complementary copy of the allele-1 tail.)

4. Signal generation – PCR round 3:

In further rounds of PCR, levels of allele-specific tail increase. The fluor labelled part of the FRET cassette is complementary to new tail sequences and binds, releasing the fluor from the quencher to generate a fluorescent signal.

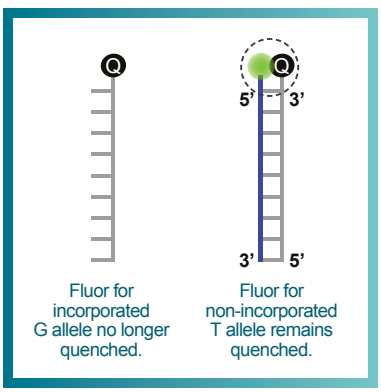


Figure 10. KASP chemistry schematic.

Detailed videos explaining the KASP reaction components, how the chemistry works and how results are analysed can be accessed on our website.

To run KASP in your own laboratory, you will need to submit DNA sequence for your SNP(s) of interest. Our in-house software will use the information provided to design the primers for each assay. These will be synthesised, combined and shipped to you in one tube of Assay mix for each SNP or InDel of interest. In addition, you will need to purchase the appropriate version of KASP Master Mix for the instrument that you plan to run the genotyping reactions on. More details on KASP assay design and selecting the most suitable KASP Master Mix for your laboratory can be found on our website. You can also find guides for running KASP on many different qPCR instruments, including the IntelliQube™ and Nexar™ instruments.

6.2. Appendix B: Genotyping with BHQplus probes

Genotyping with BHQplus probes requires two allele-specific BHQplus probes (one for each polymorphism) and two primers that flank the region of the polymorphism. The probes are each fluorescently labelled with a different reporter molecule, and also contain a quencher. The quencher dampens the signal of the reporter when in close proximity. BHQplus probes incorporate propyne modified bases to enhance the duplex stability, increasing the probe T_M (melting temperature) beyond that indicated by the standard bases alone, enabling the design of shorter probe sequences. The enhanced specificity makes BHQplus probes perfect for genotyping single nucleotide polymorphisms. BHQ Probe Master Mix can be used for genotyping with BHQplus probes. This master mix contains *Taq* polymerase and dNTPs, in an optimised buffer solution (ROX may also be included as a passive reference dye).

During thermal cycling, the relevant allele-specific BHQplus probe binds to the target DNA. The flanking primers also bind and, as the PCR progresses, will be extend by the polymerase. As the polymerase encounters the probe, its 5’ to 3’ exonuclease will hydrolyse the reporter molecule from the probe. The increased distance between the reporter and quencher results in the permanent generation of fluorescent signal. Bi-allelic discrimination is achieved through the competitive binding of the two allele-specific BHQplus probes.

A schematic of how end-point genotyping using BHQplus probes works is shown in Figure 11.

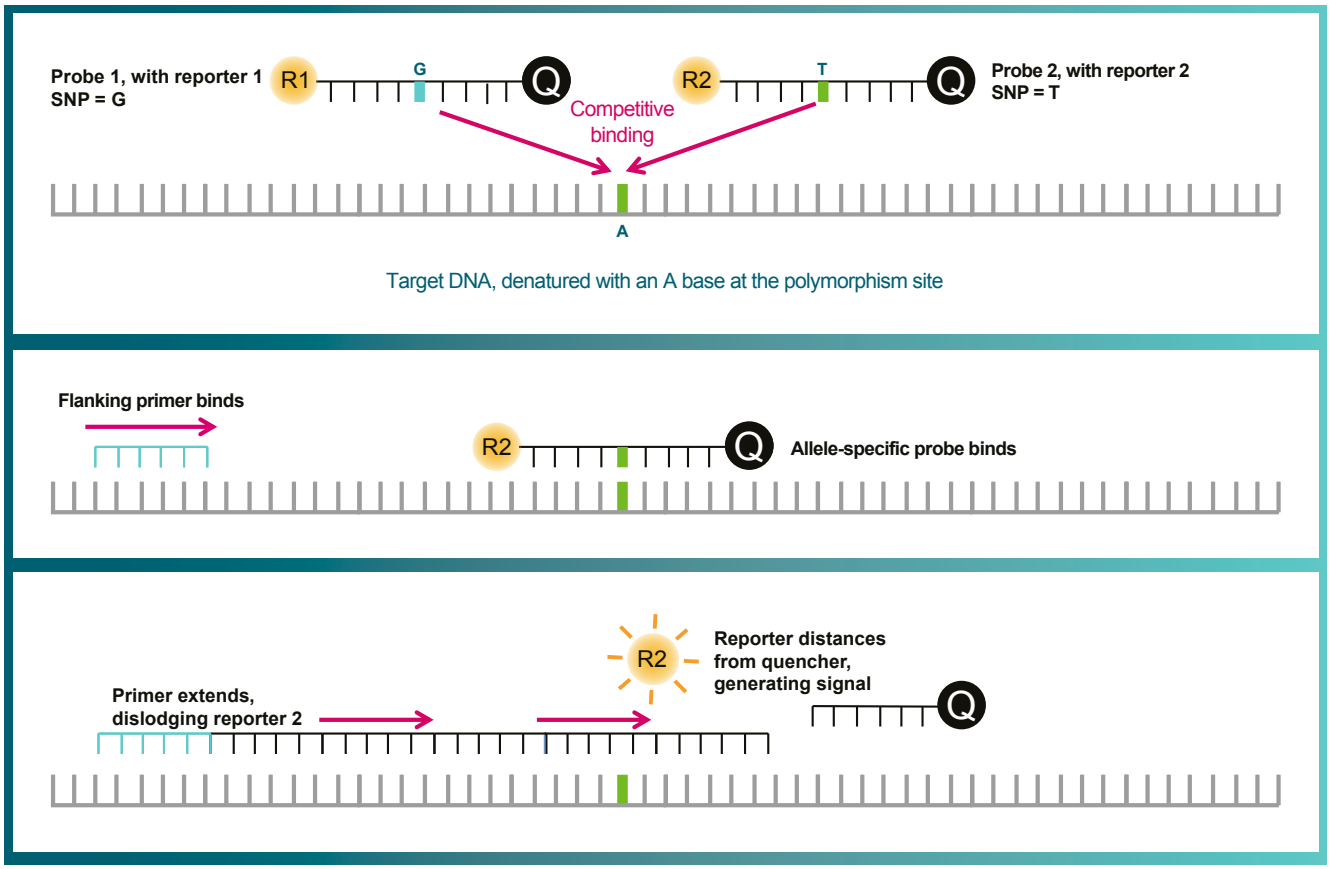


Figure 11. BHQplus probe genotyping schematic.
The polymorphism site is highlighted in green on the denatured DNA strand. The BHQplus PCR reaction mix contains the two allele-specific probes. Probe 1 has a reporter molecule that can emit a specific fluorescence, and has a G base at the polymorphic site. Probe 2 has a reporter molecule that can emit a different fluorescence to that of reporter 1, and has a T base at the polymorphic site. There is competition between probe 1 and probe 2, and probe 2 will predominate in binding as its sequence is fully complementary to the target DNA. The flanking primer binds and extends, dislodging the reporter from the probe. As a result of the increased distance between the reporter and quencher, the reporter is no longer quenched and signal is generated.

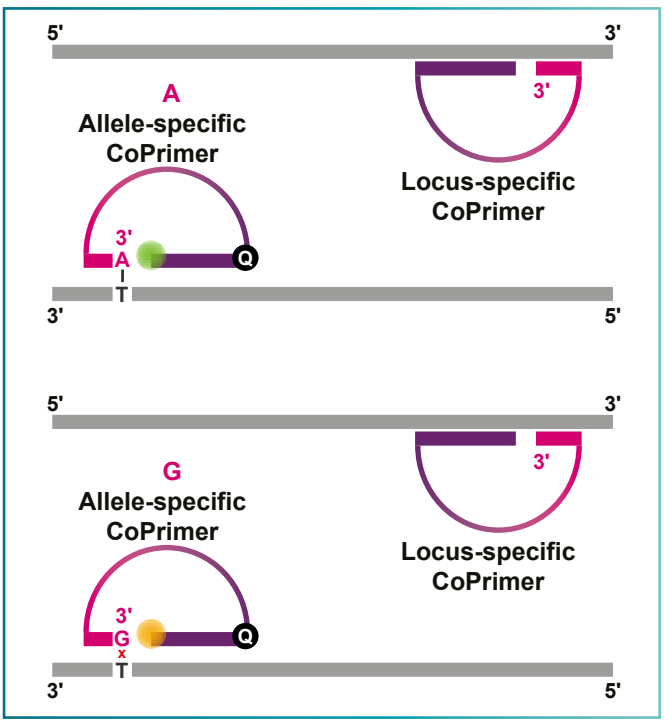
More information about BHQplus probes can be found on our website.

To run BHQplus for end-point genotyping in your own laboratory, you can either choose to order BHQplus probes and primers individually, or purchase a ValuMix Assay for SNP Genotyping which combines all of the required oligos in one tube. In both cases, primer and probe sequences will be designed by you using Biosearch Technologies' RealTimeDesign™ (RTD) software. This can be accessed free-of-charge via our website. In addition, you can purchase the appropriate BHQ Probe Master Mix for the instrument that you intend to run your genotyping reactions on.

6.3. Appendix C: Genotyping with BHQplex CoPrimers

In an allelic-discrimination genotyping reaction, a BHQplex CoPrimers assay will consist of a single reverse primer and two allele-specific BHQplex CoPrimers. Depending on the polymorphism present in the target DNA, the correct allele-specific BHQplex CoPrimers will compete for specific hybridisation.

BHQplex CoPrimers are fluorescently-labelled probes that consist of a short “Primer” sequence and a “Capture” sequence, connected by a non-extendable flexible linker to enable cooperative binding to the target of interest. They also contain an internal BHQ dye (quencher), which greatly reduces the fluorescence emitted when in proximity to the reporter dye. The components of a BHQplex CoPrimer reaction are shown in Figure 12.



Legend	
A / G	Allele-specific priming agents (forward primer)
	Locus-specific priming segment (reverse primer)
	Capture segments (forward and reverse primers)
	Flexible linker
	Fluorophores
Q	Quencher
→	Primer extension

Figure 12. Components of a BHQplex CoPrimer reaction.

On extension of the Priming fragment, the fluorescent dye is cleaved, releasing the fluorophore, and generating a fluorescent signal. The non-extendable flexible linker connecting the Primer and Capture sequences prevents the polymerase from extending through the Capture sequence, retaining the primer specificity in each round of amplification.

The Primer and Capture sequences must both correctly bind to the target in order for the polymerase to extend the Priming sequence. The non-extendable linker blocks extension of the Capture sequence alone when complementary Primer sequence does not have target-specific match. Bi-allelic discrimination is achieved through the competitive binding of the two allele-specific BHQplex CoPrimers.

A schematic showing reaction progression of a BHQplex CoPrimers assay is shown in Figure 13.

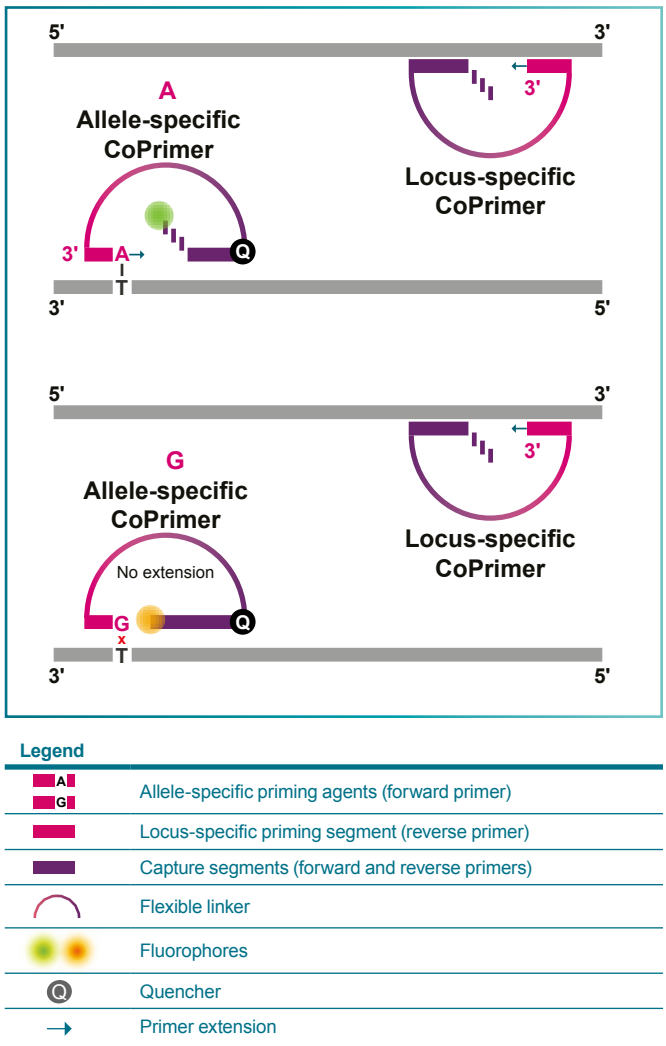


Figure 13. The reaction progression of a BHQplex CoPrimers assay.

User guide

Analysis of end-point genotyping data using cluster plots

BHQplex CoPrimers is a multiplexing PCR technology that enables assay interchangeability from a set of functional designed assays without introducing cross-reactivity. This means that you can generate twice as much data per well.

To run BHQplex CoPrimers for end-point genotyping in your own laboratory, you will need to submit DNA sequence for your SNP(s) of interest. In addition, we will need details of the dye combinations and the oligonucleotide synthesis scales that you require. Our spectral overlay tool and our dye selection chart can be used for assistance in selecting appropriate dye combinations for multiplexing. The BHQplex CoPrimer Design software suite will be used to design the assays. These will be synthesised and shipped to you as lyophilised individual oligonucleotides. In addition, you will need to purchase the appropriate version of BHQ Probe Master Mix, which is the recommended PCR Master Mix to use to maximise the performance of your BHQplex CoPrimer genotyping assays.

Integrated tools.
Accelerated science.

@LGCBiosearch | biosearchtech.com

BIOSEARCH™ TECHNOLOGIES

GENOMIC ANALYSIS BY LGC