APPLICATION INFORMATION

Genetic Analysis

AUTOMATING HETEROZYGOTE DETECTION USING THE HUMAN P53 GENE AS A MODEL SYSTEM

Mark Dobbs and Nancy Gee
Beckman Coulter, Inc.

Identifying heterozygous base positions in genomic DNA can be of great utility in locating mutations or polymorphisms in diploid organisms. DNA sequencing is the ultimate confirmation of the presence of more than one base at a single nucleotide position, but heterozygous base positions are often overlooked in sequencing results. The CEQ™ analysis software has been enhanced to include an algorithm that searches for and labels heterozygous base positions. The user parameters and capabilities of the enhancement are discussed below.

Method for Simulating and Testing Heterozygosity

All templates used in these experiments were PCR* products derived from a cloned human p53 PCR product. A region including exons 7 and 8 was amplified using primers 7-8a and 7-8b (Figure 1 and Table 1) and cloned into pCR†2.1-TOPO† using the Topo TA Cloning† kit from Invitrogen. PCR products from the clone were altered by site directed mutagenesis with degenerate oligonucleotide primers and re-cloned as above. Clones were sequenced, and selected PCR products from the modified clones were mixed in 50:50 ratios to simulate heterozygosity. A number of unintentionally created mutations were found outside the targeted dinucleotide region in codon 248, and these were used to validate additional heterozygosities. The full-length, 622 base pair PCR product was sequenced with each PCR primer plus two internal reverse primers (Table 2).

Twenty-five fmol of fifteen different PCR product sequencing templates were combined in eleven combinations of two. Twenty microliter sequencing reactions were performed in polypropylene 96-well plates, and the reactions were precipitated in the same plates using the standard Beckman Coulter ethanol plate precipitation protocol.[1] These combinations enabled us to interrogate 16 heterozygosities at 12 different positions along the p53 PCR product. The heterozygous base positions examined by each sequencing primer are indicated in Figure 1 and listed in listed in Table 3.

Method LFR-b was used for all separations unless otherwise noted (see the table below).

Method Name	*Injection Conditions*	*Separation Temperature (°C)*	*Separation Voltage (kV)*	*Separation Time (min.)*
LFR-a	2.0 kV, 10-15 sec	50	4.0	110
LFR-b	2.0 kV, 10-15 sec	57	6.0	60

Eight replicates were performed for each template mixture and primer. All separations in these experiments were performed on short 33 cm × 75 µm capillaries (Beckman Coulter P/N 608087). Samples were injected for 10 seconds. The LFR methods used a single ramp of 5 minutes to final separation voltage. The default heterozygote detection parameters which we used in this study are shown in Figure 2b and explained graphically in Figure 3.

Analysis Parameters for Heterozygote Detection

To initiate heterozygote detection, the user must first check the **Detect Heterozygotes** checkbox on the Sequence Analysis Parameters editor. Checking the Detect Heterozygotes option automatically deactivates the **Call Threshold function**, allowing the software to assign R, Y, K, M, S, or W codes at heterozygous base positions (see Figure 2a and Tables 4 and 5). Checking the PCR product option terminates the analysis at the end of data.

Heterozygote detection takes place after the original basecall of the sequence is complete. First a list of all peaks that meet the sensitivity criterion is constructed. Then each called base is evaluated to see if the detected peaks meet two further criteria—the **% of Average Peak Spacing** and the peak **Height Ratio**. The % of Average Peak Spacing value is used to set the size of the window around each called base. This window defines the region to look for additional peaks (where 100% is halfway between peaks). It is calculated using the average spacing of the ten nucleotides centered on the base in question. The **Height Ratio** is a minimum relative height of a lower secondary peak, expressed as a percentage ≤100%. The **Sensitivity**, a setting which assesses the slope of a peak, is adjustable from 0 to 1.0, where a higher setting will detect peaks that are very shallow in slope or peaks that may be shoulders within other peaks. The **Range** allows the region of heterozygote analysis to be restricted. The end of the range can be set as an absolute number, or as a number of bases before the last base called by the analysis software. If numbers are present in both range and relative base fields, the heterozygote analysis will proceed to whichever base number is encountered first.

The default heterozygote detection parameters which were used in this study are shown in Figure 2b and graphically explained in Figure 3.

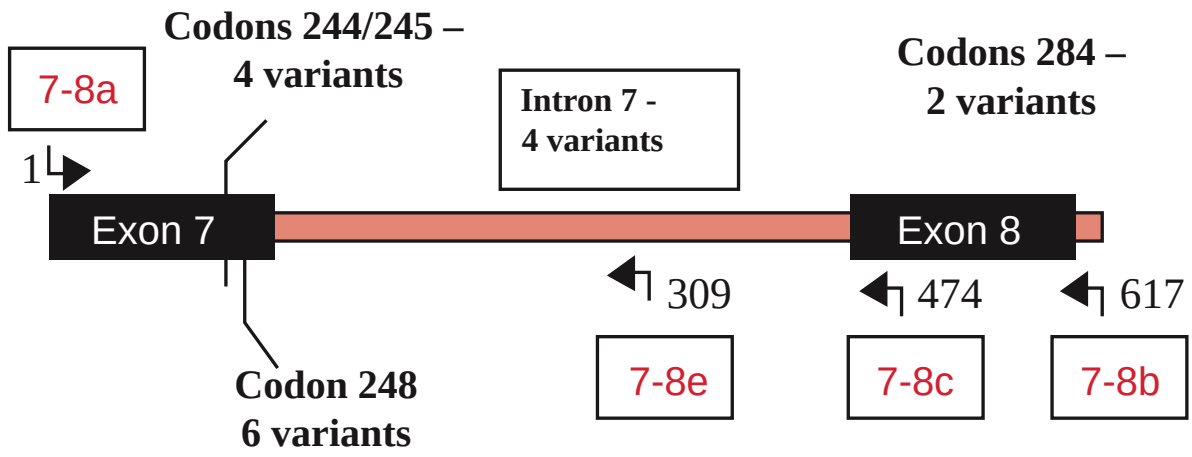


Figure 1. *Layout of the Exon 7-8 PCR product used in these experiments. Oligonucleotide primers are bent arrows labeled in red. The relative positions of two internal reverse primers (7-8c and 7-8e) as well as the end primers (7-8a and 7-8b) are indicated.*

Table 1. Base Numbering for p53 Exons 7 and 8

The sequence was obtained from Genbank using sequence accession number X54156.
The 3´ end of primer 7-8a was within the coding region of exon 7.

Gene Element	*Starting Base Number*	*Ending Base Number*	*Number nt's*	*Number Encoded aa's*	*Codon Numbers*
Exon 7	14,000	14,107	108	36	225-260
Intron 7	14,110	14,451	342		
Exon 8	14,450	14,587	138	46	261-306

Results and Discussion

The heterozygote analysis module proved capable of detecting all expected heterozygosities in our collection of p53 mutations. These heterozygosities represented each combination of two bases in a variety of sequence contexts providing a good model to test the detection algorithm. We were even able to accurately and reproducibly assign heterozygosity at positions more than 500 bases into the called sequence.

Depending on the heterozygosity analyzed, the ratio of the larger to the smaller peak was variable, sometimes being 1:1, and never greater than 2.5:1. This variability in peak height ratios is a function of the different incorporation rates of the four dye-terminators used. The variability, however, had no effect on the algorithm's ability to accurately sequence or identify heterozygotes. The relative peak heights of the two bases at each position of heterozygosity was constant from sample to sample (see example in Figure 4) highlighting the precision of the dye terminator cycle sequencing chemistry. In this experiment, a **Height Ratio** of 30% was more than enough to capture the lowest relative peak heights that we observed.

Table 2. Primer Sequences

The primers were calculated to have Tms between 65 and 70°C.

Primer Name	*Orientation*	*Sequence*
7-8a	Forward	5´ GGA ATT CTC CTA GGT TGG CTC TGA C 3´
7-8b	Reverse	5´ GGA ATT CCT GCT TGC TTA CCT CGC T 3´
7-8c	Reverse	5´ CAA ACA CGC ACC TCA AAG CTG TTC C 3´
7-8e	Reverse	5´ GCT CCA GGT AGG TGG AGG AGA AGC C 3´

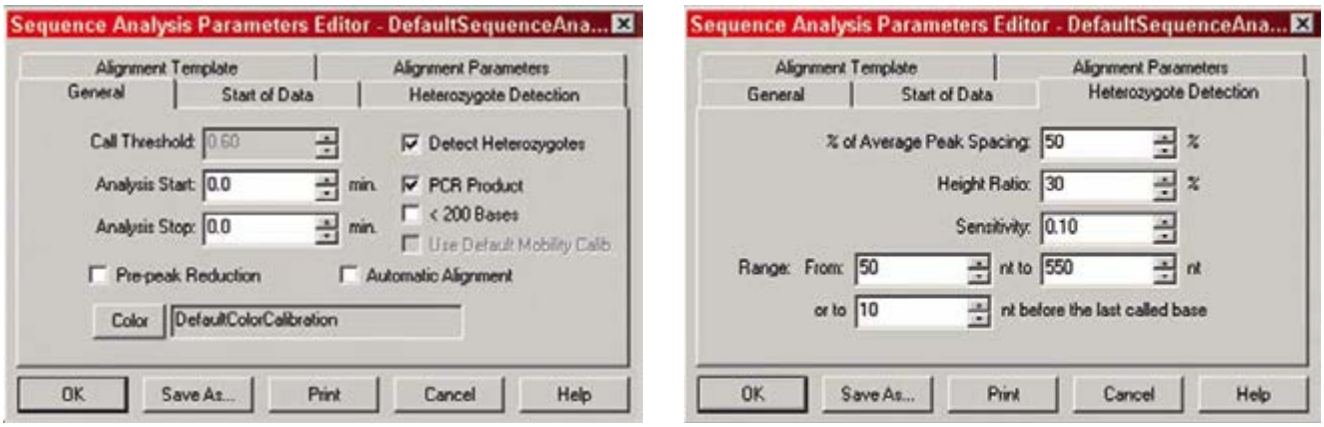


Figure 2. *User-defined variables for the detection of heterozygous base positions.*

a) *General analysis parameters.* ***b)*** *Heterozygote detection parameters.*

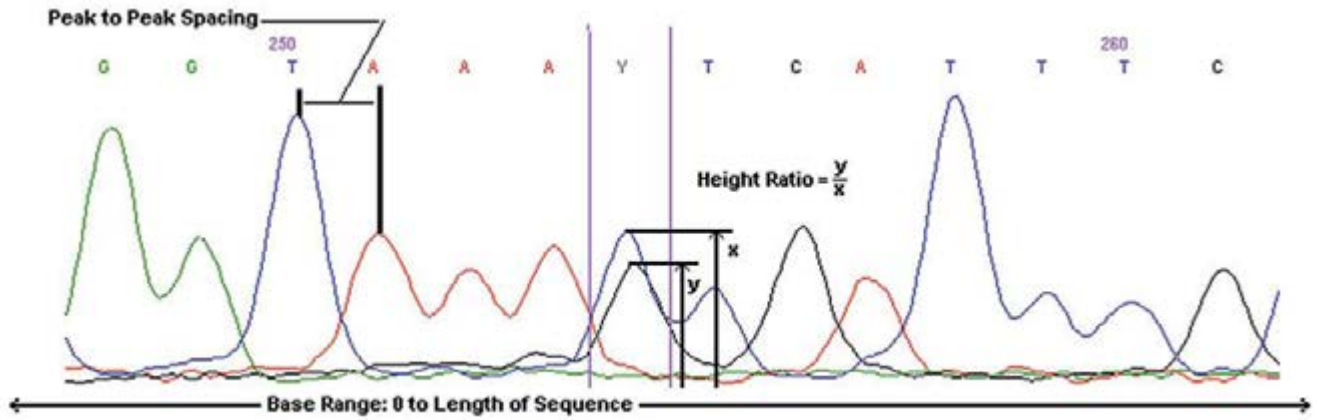


Figure 3. *Graphical representation of heterozygosity and user-settable variables.*

Although most of the heterozygosities were immediately detected, a small number (6/84) did require some adjustment to the default detection parameters. This occurred when:

1) At least one of the two peak centers was outside a window that was 50% of the expected peak to peak spacing (25% to either side of the expected peak center)
2) One of the two peaks resided in the shoulder of an identical neighboring peak

In this data set, there were three examples where a peak spacing setting between 70% and 85% was required to recognize the heterozygosity. Two of the heterozygotes occurred between 50 and 70 bases in the called data (Figure 5). The aberrant peak spacing was readily visualized in both the electropherogram and the quality value plot, as illustrated by a closer than normal peak spacing and a slight dip in quality values around the heterozygous base position. The third case occurred at base 525 with primer 7-8b (Figure 6). By simply increas-

Table 3. List of p53 Mutations Created in Site-Directed Mutagenesis Experiments

Exon/ Intron	*Nucleotide*	*Codon*	*WT*	*Mutants Available*	*AA Change*
Exon 7	14057	244	G		
	14058	244	G		
	14059	244	C	T	G>G (Silent)
	14060	245	G	C	G>R
	14061	245	G	C	G>A
	14062	245	C		
	14063	246	A		
	14064	246	T		
	14065	246	G		
	14066	247	A		
	14067	247	A		
	14068	247	C		
	14069	248	C	A,G,T	R>R (Silent),G,W
	14070	248	G	A,T	R>Q, L
	14071	248	G		
Intron 7	14194		G	A	NON-CODING
	14239		G	A	NON-CODING
	14428		C	T	NON-CODING
Exon 8	14519	284	A		
	14520	284	C		
	14521	284	A	G	T>T (Silent)

Table 4. One-Letter IUB Codes for Heterozygous Base Positions

Bases	*One-Letter Base Code*	*Opposite Strand Base Code*	*Mnemonic*
A or G	R	Y	pu**R**ine
C or T	Y	R	p**Y**rimidine
G or T	K	M	**K**eto
A or C	M	K	a**M**ino
G or C	S	S	**S**trong
A or T	W	W	**W**eak

ing the % of the Average Peak spacing from 50% to 70% the call were accurately made.

There is relatively little risk in raising the **% Average Peak Spacing** above 50% for routine heterozygote detection, as long as the range for detection does not include irregularly spaced bases at the beginning of a sequencing result. Uniform spacing early in the sequence depends on good sample cleanup. In these experiments, neither the base spacing adjustment nor the sensitivity adjustments created any false heterozygous calls.

High sensitivity detection of heterozygotes is demonstrated in Figure 7, where the first G of a GG doublet, that was barely visible as a shoulder of the following peak, was correctly identified as a heterozygote. By simply raising the **Sensitivity** value to 0.75, the software made the correct assignment of the heterozygous base. However, it should be noted that with such a high sensitivity setting, secondary sequencing products, such as those associated with contaminating templates or primers, could be identified as heterozygotes. It is important to consider the purity of the sample preparation when optimizing the heterozygote detection software settings.

While nearly all of the heterozygosities were detected using the faster LFR-b separation method, some heterozygosities that occurred after 450 bases were more easily detected when using the LFR-a method (see Figures. 6 and 7). This method appears to be the best method for detecting heterozygosites that occur after 450 bases into the data, since it provides better resolution of sequencing peaks further into the separation.

Overall we found advantages to sequencing both strands of the DNA template when evaluating a sequence for heterozyosity. While high quality DNA sequence extended well beyond 500 nucleotides, sequencing from the opposite end of the DNA

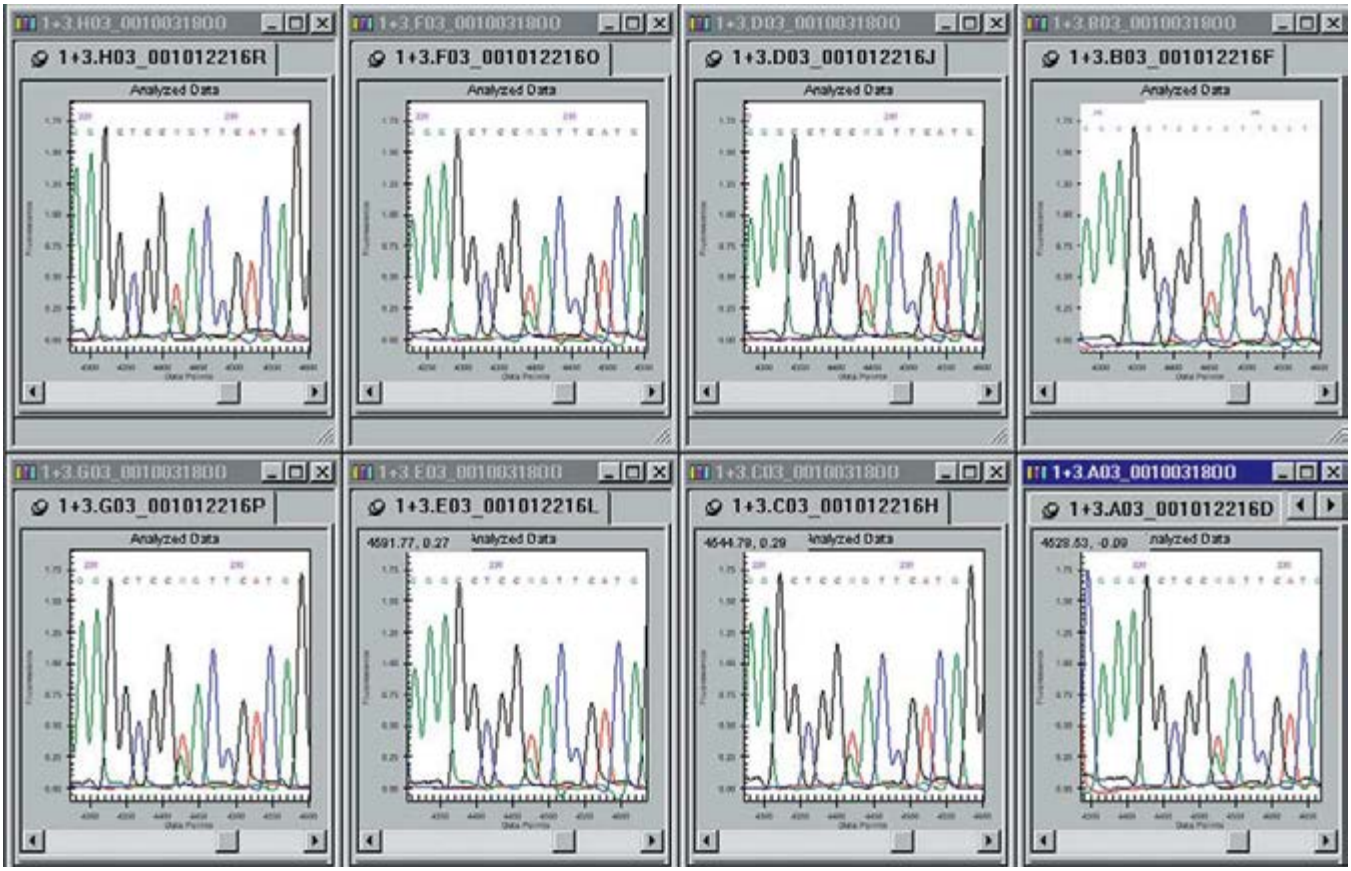


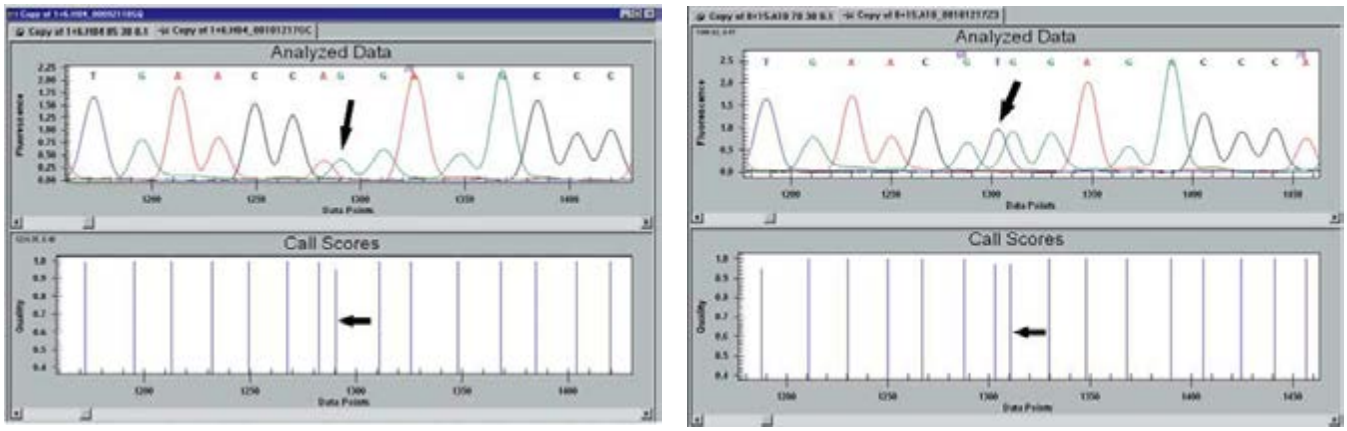
Figure 4. *An example of the reproducibility of heterozygote peak heights and peak spacing for replicate sequencing reactions. The template is a mixture of wild-type PCR product and mutant C14069T. Using the reverse prime 7-8e, the heterozygosity is visualized as G (green) and A (red) peaks at approximate base position 225 of the called sequence.*

template inward compensated for the gradual decline in resolution that is unavoidable in DNA sequencing reactions. In addition, since the sequence context surrounding any one base position is different on the two strands, all of the context-specific peak height variations on one strand were compensated for by non-synonymous variations on the other strand.

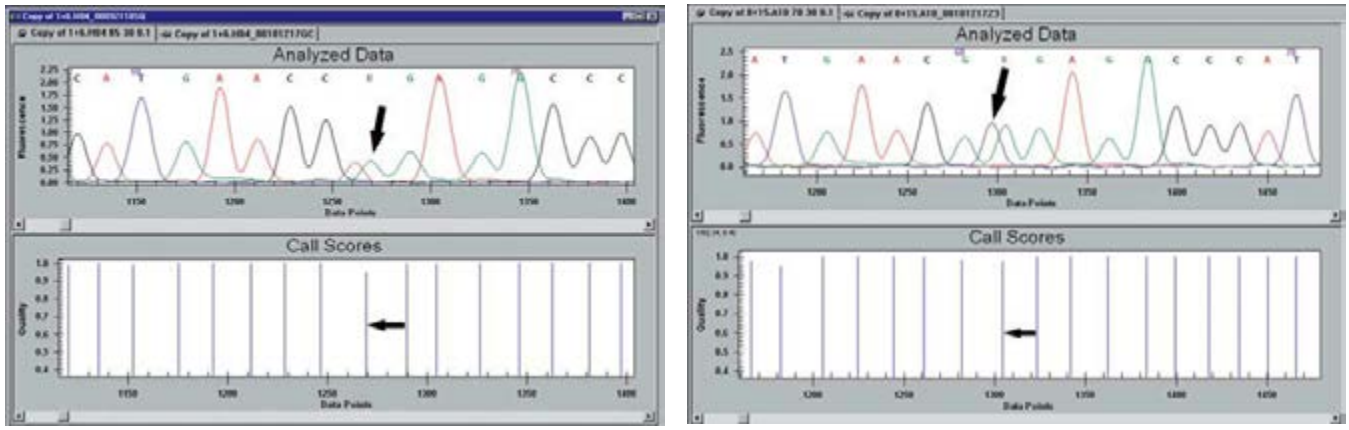
Reference

1. *CEQ™ 2000 Dye Terminator Cycle Sequencing: Chemistry Protocol: Step-by-Step Guide*, Beckman Coulter P/N 718119B. http://www.coulter.com/Beckman/biorsrch/BioLit/Pubs/718119ab.pdf

Figure 5. *Two of 84 scored heterozygosities had peak centers that were slightly outside the default peak spacing window.*



a) *Default heterozygote detection parameters. Quality Value plots indicate an abnormality in the sequence. Note the closer than normal spacing of the quality value bars, and the slight dip in quality values compared to neighboring values indicates a potential uncalled heterozygosity.*



b) % Average peak spacing *raised to 85% or 70% respectively. The quality value bars are more evenly spaced when the heterozygous base positions are properly identified.*

Table 5. Combined Number of Each Class of Heterozygosity Examined

Except for S and W, the heterozygosity code at any one position will change when the opposite strand is sequenced. The exon 8 heterozygosity was late in the 7-8a DNA sequence, and therefore was not scored. The 7-8c and 7-8e primers excluded one or two heterozygosities that were present in the 7-8b sequence.

Primer	*Combined Number of Each Class of Heterozygosity Examined*						
	R	***Y***	***K***	***M***	***S***	***W***	***Total***
7-8a (forward)	5	8	2	1	4	1	21
7-8b (reverse)	8	6	1	2	4	1	22
7-8c (reverse)	8	5	1	2	4	1	21
7-8e (reverse)	7	5	1	2	4	1	20
Total	28	24	5	7	16	4	84

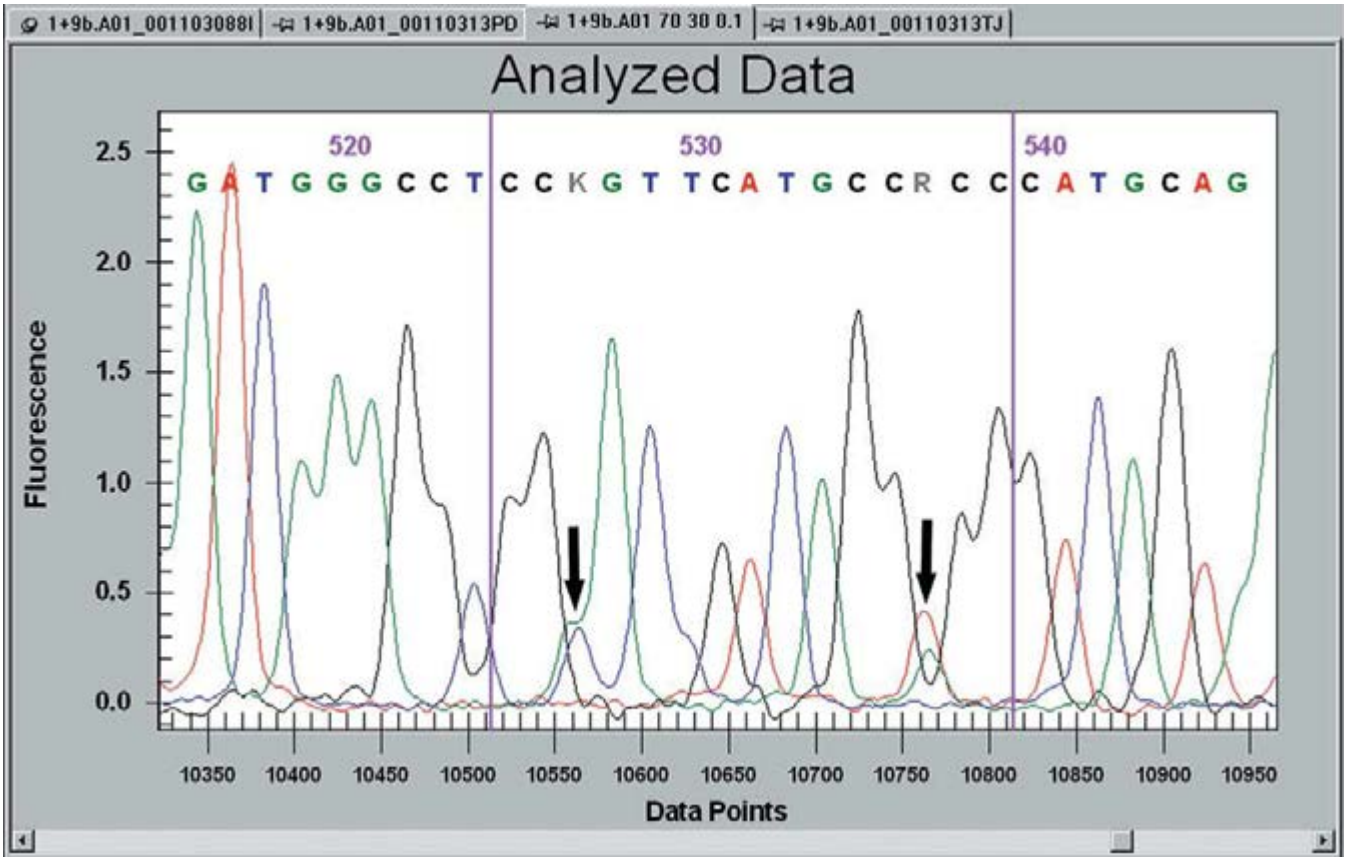


Figure 6. *Example of a heterozygosity that was picked up with a* ***% of Average Peak Spacing*** *70% versus the default % of* ***Average Peak Spacing*** *of 50%. Note the second heterozygosity at position 536. The separation method used for this sample was LFR-a.*

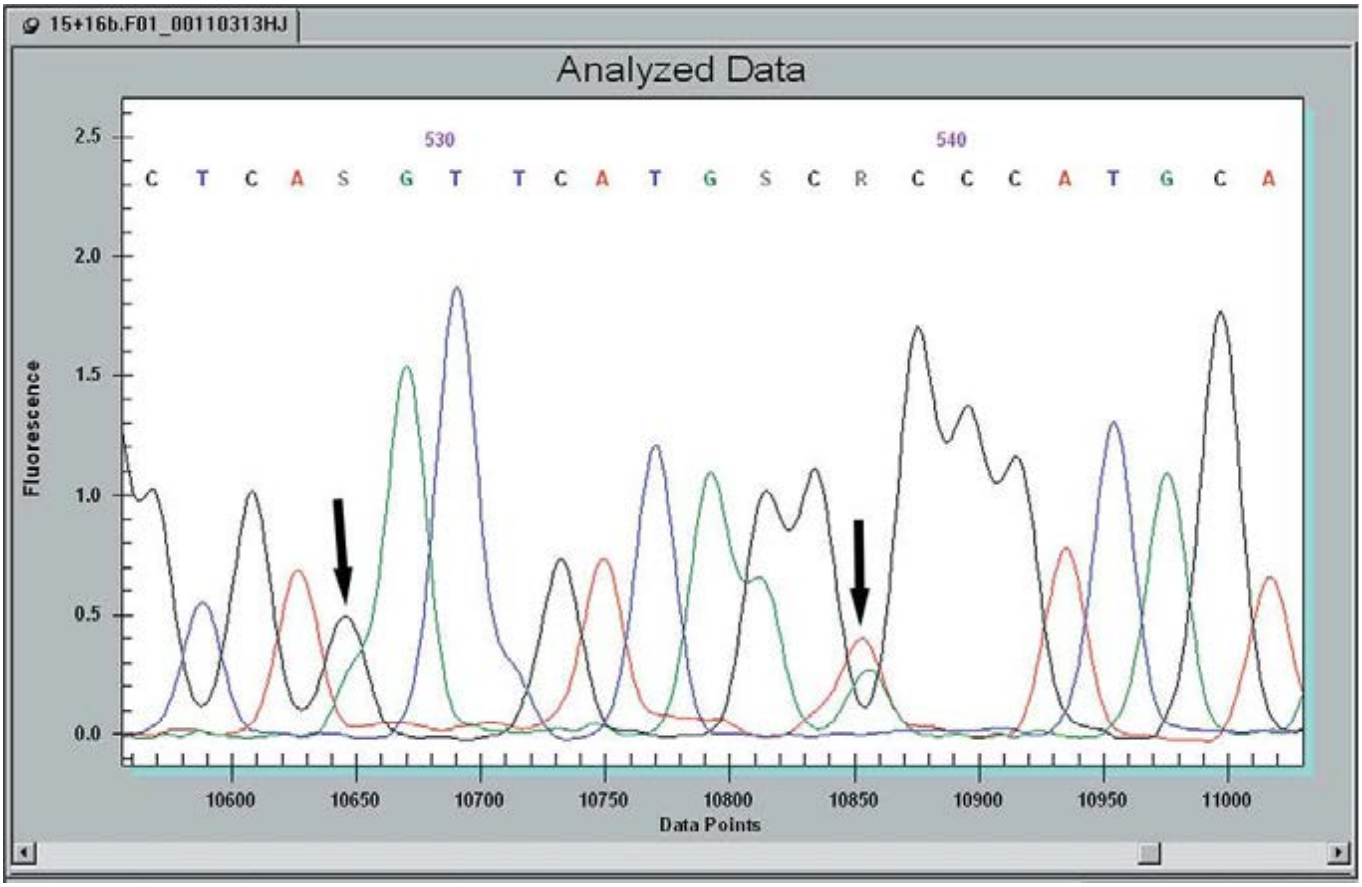


Figure 7. *The heterozygosities at base positions 536 and 538 were made under default heterozygote detection settings. The heterozygosity at position 528 required increasing the sensitivity from 0.1 to 0.75. The separation method used for this sample was LFR-a.*

* *PCR is covered by U.S. patents owned by Hoffmann-La Roche, Inc.*

† *Topo is a trademark and pCR and TA Cloning are registered trademarks of Invitrogen. All other trademarks are the property of their respective owners.*

I.S. EN ISO 9001

www.sciex.com/ce

View SCIEX products at www.sciex.com
Find your local office at www.sciex.com/offices

AB SCIEX Headquarters
500 Old Connecticut Path | Framingham, MA 01701 USA
Phone 508-383-7700
www.absciex.com