

## Genetic Analysis

# SNPs. MUTATIONS AND DNA SEQUENCE VARIATION ANALYSIS USING THE GENOMELAB™ SNPSTART KIT

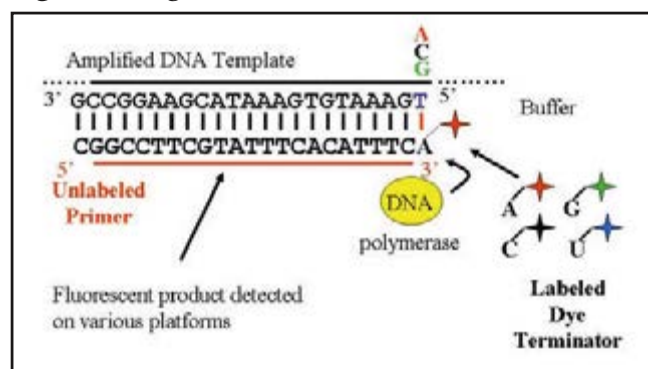
Nitin Udar, Jana Mariana, Margaret Porter and Doni Clark  
Beckman Coulter Inc., Fullerton, CA, USA

### Introduction

SNP analysis is routinely used for linkage, association studies, pharmacogenetics, forensics, etc. Similarly, mutation screening requires detection of single base changes, as well as analysis of insertions, deletions and translocations among other molecular events. In this study we have used the single-base extension (SBE) technology to extend an unlabeled primer flanking a target site with a fluorescent labeled terminator.

The product is separated and detected on a capillary electrophoresis platform – Beckman Coulter CEQ™ 8000 Series Genetic Analysis System. All the reagents necessary for the SBE reaction including the fluorescent dyes, enzyme and buffer, are bundled in a single ready-to-run SNPStart master mix (*Figure 1*). The master mix is optimized to generate accurate as well as well-balanced allele signals in a SNP reaction. We demonstrate multiple levels of plexing capability using SNP panels from one to 10-plex. In addition, multiplex Cystic Fibrosis (CF) mutation panels were also tested using patient DNA samples. Signal balance between two heterozygous alleles is important for accurate detection. We measured this signal balance for each of the two alleles of the six possible heterozygote combinations and found an average signal ratio of 1:1.6. The allele calling accuracy for a 10-plex reaction is >99%. Comparisons were done using different sources of DNA, including DNA extracted from cell lines and prepared from whole genome amplification. These were used as templates for PCR and subsequently used for primer extension reactions. Our results demonstrate an efficient process of producing more than 960 SNP/mutation screens per plate. DNA extracted on the Vidiera™ NsP nucleic acid sample preparation platform (based upon the Biomek® NX

*Figure 1: Single-Base Extension Reaction*



platform) was tested for the SNP application. This testing was carried out with the intent of increasing process performance and reducing possible sampling errors and hands on time per sample. In this report we demonstrate the capabilities and ease of use of the SNPStart Primer Extension kit.

### Methods and Materials

#### PCR Primer Design

Primers for PCR were designed using the Primer 3 software developed by the Whitehead Institute for Biomedical Research. PCR primers were designed with melting temperature between 60°C - 80°C with optimum between 70°C - 75°C. The length of the primer was designed between 23-32 nt with an optimum of 27 nt. Avoid designing PCR primers in repeat regions.

Now sold through SCIEX Separations  
[www.sciex.com/ce](http://www.sciex.com/ce)

SCIEX

BECKMAN  
COULTER

Capillary Electrophoresis

The desired length of the PCR products was between 150-1000 bps. The lower limit is important when columns are used for PCR purification. Most commonly used columns have a lower limit of approximately 100 bp. PCR products larger than 1000 bps may need an additional denaturation step (94°C for 1 min) immediately prior to primer extension cycling. Smaller PCR products are desired because they will have a less likelihood of nonspecific binding of other products. For a multiplex PCR situation, design the product lengths so that they do not overlap by less than 30bp. This enables visualization of each individual band if run on a regular agarose gel to verify successful PCR before further processing the sample. In these experiments we used the following PCR cycling conditions: 94°C/2 min followed by 35 cycles of 94°C/30 sec, 60°C for 1 min and 72°C for 1 min (the time will need to be increased further for longer PCR products) and then held at 4°C. The PCR products have to be further processed to remove all the excess primer and dNTP before proceeding with the SNP reaction. This can be achieved by either treating the PCR product with EXO-SAP (GE HealthCare, US) or column purification. Follow manufacturer's recommendation for this step. PCR products were electrophoresed either on an agarose gel or on an Agilent Labchip (Agilent Technologies, US) to verify specificity and make sure there were no artifact bands.

Allele balance is partially dependent on the amount of template used in the PCR reaction. If the starting DNA concentration is too little it will lead to a successful PCR reaction but the allele balance could be unacceptable. To avoid this problem of allele bias introduced during PCR, add sufficient template DNA (in the range of 100-200 ng).

### Interrogation Primers

Interrogation primers were designed using the same considerations as mentioned above (for PCR primer design). Interrogation primers can sometimes be difficult to design using the ideal parameters because the locus is fixed. There are only two choices in designing the primers – either on the forward strand or the reverse strand. Vary the number of bases in the primer to increase the T<sub>m</sub>. Hairpin and other secondary structures are difficult to avoid in the interrogation primer. Select the one that has the least secondary structure. It is sometimes easier to design primer on both strands of the SNP and selecting the one that works the best. In a multiplex panel, a poly A tail should be added to separate the different SNPs. The poly A tail is added to the 3' end of the primer. In a panel with more than 7 primers, the interrogation primers should be spaced out by 6 nts. For panels 7-plex and less, space out the primer by at least 8 nt. Interrogation primer can be designed on either strand of the target sequence. Even if primers are spaced out at fixed intervals expect a few variations in the mobility.

**Table 1:** Details of interrogation primers used in SNP panels.

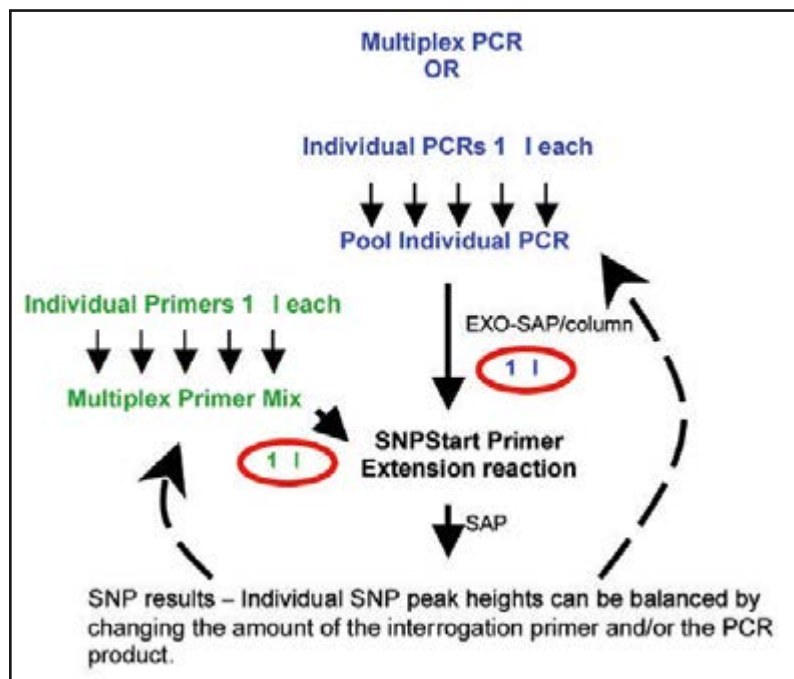
Primer	Size	Base Change	Difference in Size	Sequence
SNP1-U43	25	G/T	N.A.	AAAAGTGAAAAC TAAGGGAAACTGC
SNP2-U36	31	A/T	6	AAAAAACTGCGGGATTCTCGTTGGCAATGAA
SNP3-U13	37	C/G	6	AAAAAAAAAAAAAGTGTTTGATCTTGTGAATAAATGAG
SNP4-U19	43	C/A	6	AAAAAAAAAAAAAAAAAAAAACCCATGACTTCTGCCCCACACCCTG
SNP5-U30	49	C/T	6	AAAAAAAAAAAAAAAAAAAAAAAAAAAAATCTCAAAGGAAACACTT-TAGCCCA
SNP6-U30	55	G/A	6	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGAGATGCAACATCAGGTGATAAG
SNP7-U4	61	C/A	6	AATTTTGAACCTCCTAACAGAGCTGA
SNP8-U5	67	G/A	6	AAA-GACAGCCTCGAGACTCAGGTGA
SNP9-U34	72	C/G	5	AAA AATCTAATGGCATGTGATACTCATCC
SNP10-U37	77	C/T	5	AAA AAAAACTCTGTGGACCACATCTGAGTCCTG

\*N.A. - Not Applicable

**Table 2:** Details of interrogation primers used in Cystic Fibrosis panels.

Primer	Size	Base Change	Sequence
A455E	23	g>t	ctgctccagtggatccagcaacc
G1349D	35	g>a	AAAAAAAAAAAAAAtgggggctgtgcctaagccatg
N1303K	47	g>c	AAAAAAAAAAAAAAAAAAAAAAtgatcactccactgttcatagggatccaa
R1162X	53	g>a	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAtgttggcatgtcaatgaacttaaagactc
2789+5G	59	c>t	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAacacaatctacacaataggacatggaata
W1282X	64	g>a	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAtgttgggattcaataactttgcaacagtgtg
R117H	73	g>a	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAActtctatgaccggataacaaggagggaac
F508	37	c>t	AAAAAAAAtgcctggcaccattaagaaaatatcat
3849+10kb	31	c>a	gagtcttccatctgttcagattataaaatgg

**Figure 2:** Multiplex Setup



Generally you would require a few optimization steps to balance the signal intensity in a multiplex SNP reaction. The protocol to follow is as follows: Carry out the multiplex or single plex PCR as per your protocol. Separate it on a agarose gel to make sure there are no artifact bands (overload the product a little bit to be sure you can visualize the artifacts if any). Make sure all excess primer is removed from this reaction before proceeding with the SNP primer extension reaction. A clean and robust PCR product is

the single most important factor in determining the success of the SNP reaction. Resuspend interrogation primers to a final concentration of 100µM. Mix 1µl of each of the individual primers in a tube and use 1µl of this mixture for the SNP reaction (Figure 2). Follow the same rule of thumb for the PCR products if doing a poolplex (where each of the individual PCR products after ExoSAP/Column purification is mixed in a 1:1 proportion – take 1µl of this mixture for the SNP primer extension reaction).

### SNP Reaction

Interrogation primers used in the SNP panel are as mentioned in Table 1 and for Cystic Fibrosis Panel in Table 2. Template concentration in the various SNP reactions are shown in Table 3. If the PCR reaction required DMSO or other denaturing agents

in the reaction – purify the PCR reaction using column purification to remove the denaturing agent. If using such PCR products for the SNP reaction add Betaine in the range of 0.5 - 1.5 M to obtain robust primer extension products. The protocol as described in Table 4 was followed for the SNP and Cystic Fibrosis panels described in this report. It is important to note that the master mix has glycerol as one of its components. Therefore pipette the master mix with care. Protect the master mix as well as the reaction tubes from exposure to light.

**Table 3:** Concentration of template and primer used in the SNP reaction.

SNP reaction	Individual PCR-Template concentration in the mixture	Individual Primer concentration in the mixture	Dilutions	Loading volume
10-plex	0.3 - 3.5 ngs	0.4 - 2.0 $\mu$ M	1:3	1.0 $\mu$ l
2-plex	1.4 ngs	0.1 - 1.0 $\mu$ M	1:6	1.0 $\mu$ l
Single-plex	7.0 -40.0 ngs	0.1 - 1.0 $\mu$ M	1:10 - 1:20	1.0 $\mu$ l

**Table 4**

Component	Volume
SNPStart Master Mix	4.0 $\mu$ L
DNase RNase free PCR grade water	x.x $\mu$ L
Interrogation Primer (0.1 - 1.0 $\mu$ M)	x.x $\mu$ L
PCR product or template (1-100 fmoles)	x.x $\mu$ L
<b>Total Volume</b>	<b>10. 0 <math>\mu</math>L</b>

After mixing all components of the SNP reaction the tubes were given a short spin to consolidate the mixture to the bottom of the tube. This was followed by thermal cycling the tubes at 94°C for 20 seconds followed by 45°C for 10 seconds. The single-plex reactions were cycled 25 times and multiplex reactions 36 times followed by holding at 4°C.

After thermal cycling, the excess dye terminators in the SNP reaction were cleaned up using 0.25 U of Shrimp Alkaline Phosphatase (SAP), 1 $\mu$ l of SAP buffer and 1.75  $\mu$ l of DNase RNase free water. Load 1 $\mu$ l of the treated SNP reaction into 39.5 $\mu$ l of SLS and 0.5 $\mu$ l of Size Standard 80 (Table 4). Overlay the samples with mineral oil. After loading the sample plate and buffer plate on to the GenomeLab/CEQ Genetic Analysis System, select SNP-1 as the run method in the sample plate layout. Before starting the run make sure the capillaries are preheated to 50°C immediately prior to starting the run. This ensures proper separation for the first row of samples. Preheating is not required for subsequent rows within the same run. Observe the raw data for the first run by opening the results in the sequence analysis software. The sample peak heights should ideally be 50%-150% the peak height of size standard 13. Make appropriate dilutions of the cleaned SNP product in SLS if signal is too high. In such situations the dilution step can be avoided by decreasing the number of cycles for the primer extension reaction. Individual SNP peaks should be evenly balanced by increasing or decreasing their volume in the primer mix. In cases with a mixture of homozygous and heterozygous peaks the heterozygous peak heights should be adjusted to approximately half the height

of the homozygous peaks. If necessary use a higher (100 $\mu$ M) or lower (1 $\mu$ M) starting primer concentration for peaks that need to be adjusted. Ideally the peak heights should be between 8,000-80,000 rfu. Peaks with saturated signal will lead to errors in data analysis including missing peaks.

## Data Analysis

For Software version 8.0 and 9.0 users, install the SNP Software Update before data analysis. The software needs to be installed only once. To set up analysis parameters for the SNPStart data, use the fragment analysis module and follow the steps listed below.

1. Create a new study by selecting the sample raw data that needs to be analyzed, then select "Next".
2. Select "Default SNP Analysis Parameters" from the drop-down menu. The following parameters should be modified to correctly analyze the SNPStart data.
3. Click "Edit", then from the General Tab, change the Slope Threshold to 25 and Relative Peak Height Threshold to 15. These values can be adjusted, depending on the raw data to be analyzed.
4. Select the "Advanced" Tab and then choose the "SNP ver. 2" Dye Mobility Calibration.
5. Click "Save As", and then name the new Analysis method with a unique name, e.g. "SNPStart-Analysis Parameters." This method is now set up to analyze the data generated using the GenomeLab™ SNPStart Primer Extension Kit. This method can be further modified for adjusting the threshold or if SNP locus tags are added to the study. SNP locus tags can be created by clicking on the "SNP Locus Tag" tab in the analysis parameters section. Select "New locus" and fill in the relevant information. The text entered for the "Locus tag" will appear in the locus tag list. The text entered for the "Locus name" will appear in the analyzed results above the appropriate allele peaks. The "Apparent Fragment Size" should be determined by taking the average peak of previously analyzed results. Remove alleles that are not present in the SNP in "Allele ID's for" section. For example in an A/T SNP remove the C and G allele by deleting them. Save these settings with a different name.

Once created, the new analysis parameters will be available for subsequent projects, provided that they are in the “default” project within the same database. The parameters can also be selected for automatic analysis following each run. For frequently used parameter sets, copy and paste the new analysis parameters into the User Template Database in Data Manager to automatically populate all newly created databases with the analysis parameters.

To confirm validity of analyzed results:

- Compare the raw data with the analyzed results.
- Verify that the number of peaks (not cross-talk) is the same in raw and analyzed data. This does not take into account possible noise peaks which typically should be small.
- During a single separation, the ratio of the peak heights in the raw data within a SNP product should remain the same in the analyzed data. If there are errors in analysis due to dye matrix estimation, reanalyze the data using system dye spectra.

## Genome Amplification

The amount of DNA used for testing can become a constraint for SNP applications where a large number of different SNPs are tested or in situations where very little DNA is available. Whole genome amplification provides a method to increase the amount of starting material that can be used for downstream testing. A total of 20 ng of DNA were amplified using GenomiPhi\* DNA Amplification Kit (GE Healthcare) and used to amplify eight Cystic Fibrosis mutations. An aliquot of 0.5  $\mu$ l of the amplified product was used for the subsequent locus specific PCR reaction in a total volume of 10 $\mu$ l. After ExoSap treatment, a 0.5  $\mu$ l aliquot of the PCR product was used as template in the final SNP reaction as described above. After SAP digestion, a 1.0 $\mu$ l of the product along with 0.5 $\mu$ l of SS-80 and 38.5 $\mu$ l of SLS was aliquoted into a well. The products were separated on the CEQ 8000 system as described above. Ideally the peaks should be within 50%-150% the peak height of size standard 13.

## Automated Platform

Beckman Coulter Vidiera™ NsP Nucleic Acid Sample Preparation platform (based upon the Biomek NX platform) was used to purify genomic DNA from lymphoblastoid cell lines. This DNA was used as a template for PCR and the products obtained were subsequently used as templates for the SNP reaction.

## Results and Discussion

We carried out single-plex and multiplex SNP reactions using a panel of 10 SNPs. PCR products were obtained for each individual SNP as separate reactions. Primers were designed so that none of the products exceeded 1000 bps in size. During design they were also masked for repeats using the human repeat database. PCR products were verified on a agarose gel or Agilent chip to make sure there were no artifact bands. If artifact bands were present the PCR conditions were altered to obtain a clean single band. A clean PCR product is one of the most important criteria that determines the outcome or results of the SNP reaction. The PCR products from individual reactions for a set of predetermined heterozygous individuals were mixed together in a 1:1 ratio. Similarly SNP interrogation primers (100 $\mu$ M conc.) were also mixed in a 1:1 ratio. From each of these mixtures, a total of 1 $\mu$ l each was used for the SNP reaction. The individual template concentrations in the 10-plex template mixture were in the range of 0.3 to 3.5 ng with an average of 1.2 ng. The individual primer concentrations in the 10-plex primer mixture were in the range of 0.4 to 2.0  $\mu$ M with an average of 1.0  $\mu$ M. As illustrated in Figure 3, balanced allele signal was obtained using 3 separate SNPs with a different heterozygote combination (C/A, G/T and A/T).

During data analysis, a dye spectra is generated based on the raw data. Sometimes there is not enough data in an individual sample run for the software to calculate the necessary dye spectra. In such situations, a system dye spectra can be used to analyze the results.

Figure 4 shows a 2-plex reaction that was analyzed using a system dye spectra. Slope threshold and relative peak height threshold can be changed based on the assay to obtain fewer identification of nonspecific peaks. Figures 4 and 5 illustrate that if the slope threshold is increased from 30 to 90 and the relative peak height threshold is increased from 20 to 50, the shoulder peaks in the second C/A SNP are not called anymore.

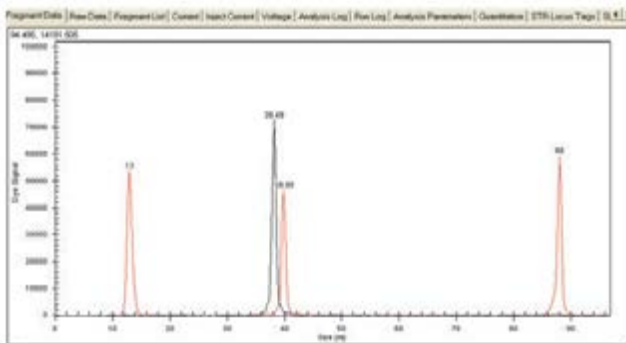
PCR products were obtained for 10 different SNP's from a set of known heterozygous DNA samples. These 10 individual products were mixed and tested using a panel of 10 SNP interrogation primers. Figure 6 illustrates the data obtained from the primer extension reaction using the SNPStart master mix.

All ten loci are heterozygous and illustrate an extreme example but in fact the allele heights for each individual peak are well balanced. Using a full plate of data (96 X 10-plex), we determined the average ratio of the signal balance to be 1:1.6 (1:1 is a ideal theoretical balance). Locus sequence characteristics influence the balance for individual SNP's. Figure 6B shows the same results but locus tags were also used during analysis. Selecting only the possible genotypes and

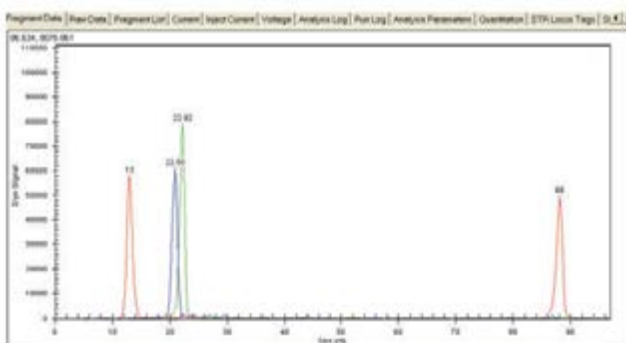
canceling the other bases is important during locus tag set up. For example, when setting up a C/A locus tag, remove the G and T options in the possible genotype. This also acts as a screening tool during allele assignment. This helps to filter out nonspecific peaks of the other colors. Use locus tags only in samples that actually will have those peaks. Unused locus tags will create unnecessary entries in the fragment list.

Similar multiplex panels were also tested to screen a panel of Cystic Fibrosis mutations. Two different panels were tested on different patient DNA samples. Figure 7 and Figure 8 shows the results obtained for these experiments. All four patients' DNA showed the correct genotype and mutations.

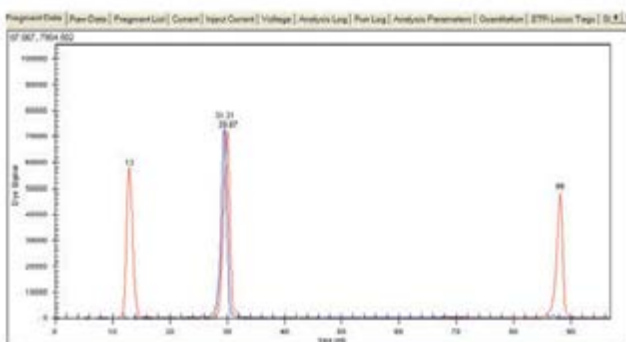
**Figure 3:** Single-plex SNP reactions - 3A) Heterozygous C/A alleles. 3B) Heterozygous G/T alleles- 1C) Heterozygous A/T alleles.



3A

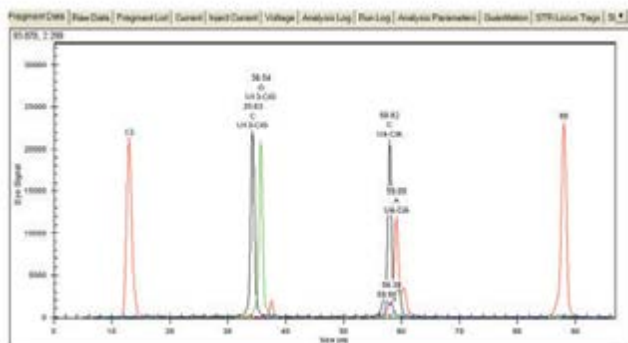


3B



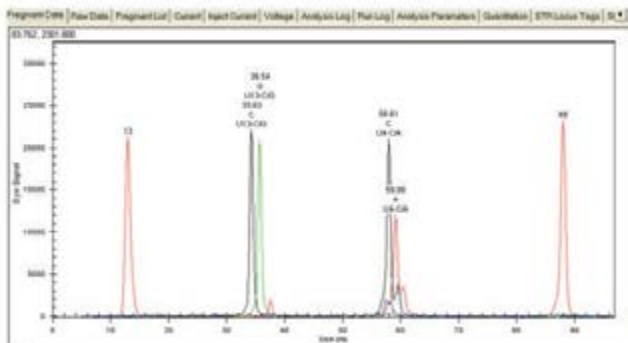
1C

**Figure 4:** Two-plex SNP reactions - both heterozygous C/G and C/A alleles.



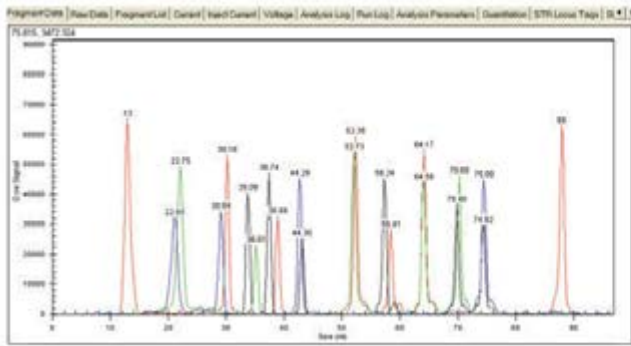
The two-plex SNP data is analyzed using A slope threshold of 30 and relative height threshold of 20. Small shoulder peaks are observed for the C/A SNP. These peaks are more commonly observed in PAGE purified primers as compared to HPLC purified interrogation primers. System Dye spectra was used to analyze the data as compared to Calculated dye spectra in order to get the correct SNP allele assignments.

**Figure 5:** Two-plex SNP reactions - both heterozygous C/G and C/A alleles.

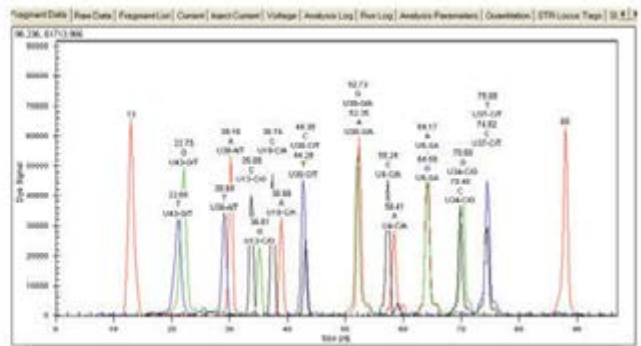


The two-plex SNP data is analyzed using A slope threshold of 90 and relative height threshold of 50. Note the shoulder peaks on the C/A SNP are not called as in Figure 4.

**Figure 6:** Ten-plex SNP reactions - All SNPs have heterozygous alleles - SNP1-G/T, SNP2-A/T, SNP3-C/G, SNP4-C/A, SNP5-C/T, SNP6-A/G, SNP7-C/A, SNP8-A/G, SNP9-C/G, SNP10-C/T. Fig 6B shows the same panel but analyzed using locus tags.

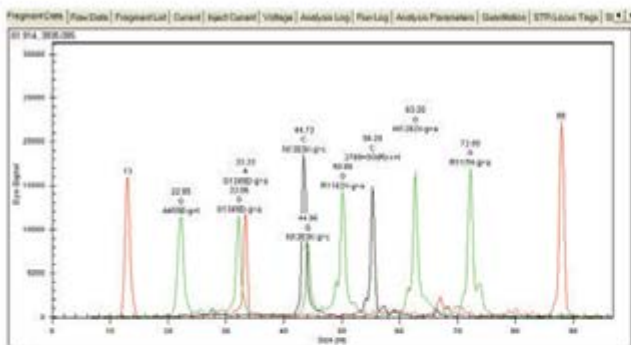


**6A**

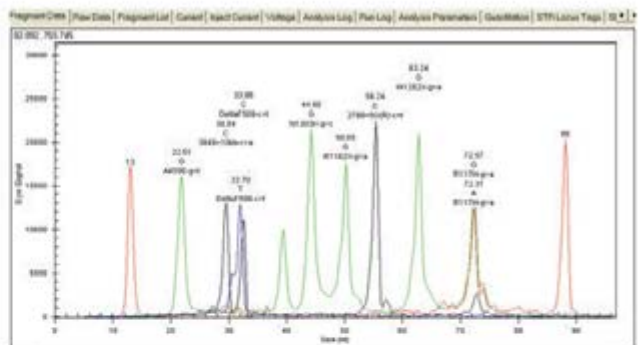


**6B**

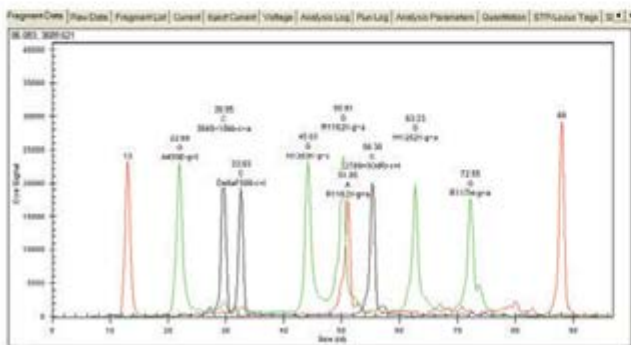
**Figure 7:** Cystic Fibrosis Mutation Panels illustrating eight mutation from patient DNA -  
DNA#12:Panel- A455E, G1349D, N1303K, R1162X, 2789+5G, W1282X, R117H  
DNA#16, #18, #21:Panel- A455E, 3849+10Kb, N1303K, DeltaF508, R1162X, 2789+5G, W1282X, R117H



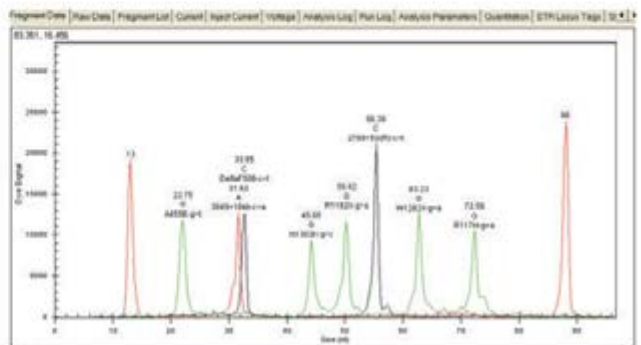
**DNA#12**



**DNA#21**

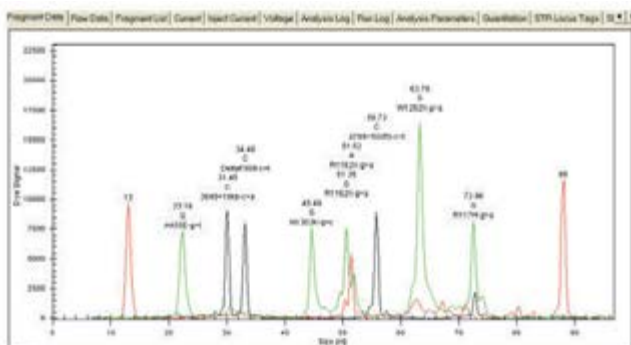


**DNA#18**

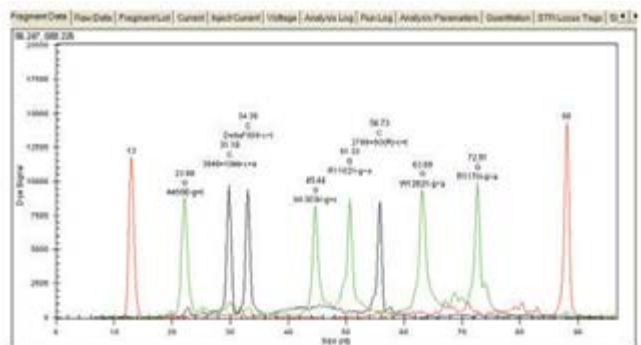


**DNA#16**

**Figure 8:** 8A is an example of a Cystic Fibrosis mutation panel tested on genome-amplified DNA as source DNA for a CF individual illustrating eight mutations one showing a heterozygous R1162X mutation and the rest are homozygous alleles. 8B shows the same eight mutations, from a normal individual where all are normal homozygous alleles.

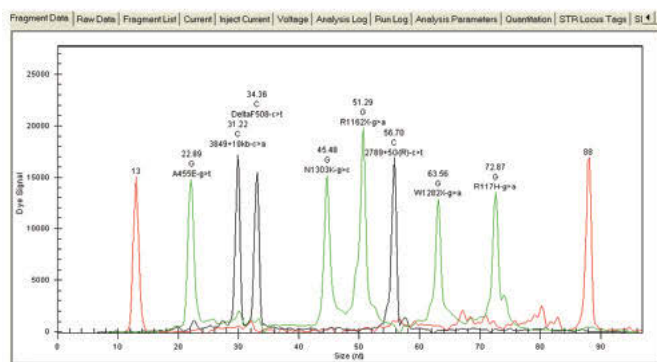


**8A**

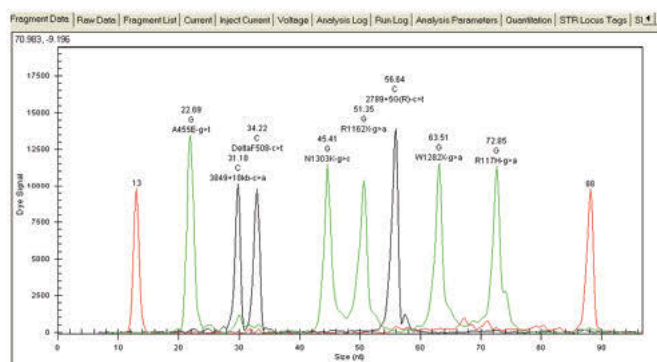


**8B**

**Figure 9:** SNPStart data for results obtained using 2 DNA samples purified on a Vidiera NsP platform (based upon the Biomek NX platform) tested for the Cystic Fibrosis Panel of mutations.



9A



9B

SNP genotyping has become a significantly powerful tool for research. SNP frequencies in the genome makes them ideal for testing for many different applications. However each individual SNP is only marginally informative (compared to microsatellite markers). To compensate for the informativeness a large number of SNPs are need to be screened. Although this may seem easy to do, the amount of starting DNA sample may make this a limiting endeavor. Genome amplification using phi polymerase offers a useful alternative. We tested genome-amplified DNA as starting material to find out if all the loci were equally amplified and if there was any difference in allele bias due the extra amplification step. Figure 8 shows that there were no significant differences in the results obtained using genome-amplified DNA for the Cystic Fibrosis locus as compared to genomic DNA isolated from lymphoblastoid cell lines.

For high throughput screening or to automate the screening process we also tested DNA samples purified from lymphoblastoid cell line using the

Vidiera™ NsP nucleic acid sample preparation platform (based upon the Biomek NX platform). DNA extracted on this platform gave robust SNP results illustrating that this platform can easily be combined to generate an automated workflow in a laboratory for accurate and efficient processing of samples (Figure 9).

## Summary

In summary, the GenomeLab™ SNPStart Kit allows users to obtain multiplex SNP data from templates generated from different sources of DNA. The color balance and allele signals are optimized in this kit. The kit comes as a ready-to-use single tube reagent that is based on a simple primer extension technology with no additional secondary steps involved. The kit has been demonstrated to run 10-plex SNP reactions. Therefore, a total of 960 genotyping results can be accomplished in a 96-well plate format.

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

© 2014 AB SCIEX. SCIEX is part of AB Sciex. The trademarks mentioned herein are the property of AB Sciex Pte. Ltd. or their respective owners. AB SCIEX™ is being used under license.