MULTIPLEX SNP ANALYSIS: SCREENING FACTOR V R506Q (LEIDEN) MUTATIONS

Zhiming Jiang[1], Milhan Telatar[2], Wayne W. Grody[2]
[1]Beckman Coulter, Inc.
[2]University of California, Los Angeles

Introduction

Single nucleotide polymorphisms (SNPs) are major contributors to genetic variation, making up approximately 80% of all known polymorphisms, and their density in the human genome is estimated to be, on average, one per 1000 base pairs. Although SNPs are mostly biallelic (less informative than short tandem repeats), they are more frequent and mutationally stable, making them suitable for association studies in which linkage disequilibrium (LD) between markers and an unknown variant is used to map disease-causing mutations. In addition, because SNPs have only two alleles, they can be genotyped by a simple yes or no assay rather than a length measurement, making them more amenable to automation.

SNP Analysis

The SNP analysis involves three stages: SNP discovery, validation, and scoring. SNP discovery involves identifying potential SNPs by technologies such as sequencing, SSCP, or DHPLC. SNP validation confirms the real mutations and their allele frequencies in different populations. The SNP scoring phase includes genetic analysis such as association or linkage. The technological and economic goal is accurate, easy, cheap, and fast large-scale SNP genotyping.

Primer Extension

The single-base primer extension is a method of sequencing the precise location of a SNP site. It utilizes the inherent accuracy of DNA polymerase to determine the presence or absence of the specific nucleotide at the SNP site. A specially synthesized DNA primer is used to anneal to the SNP site of interest. The primer anneals one base short of the target SNP. DNA polymerase inserts the complementary dideoxy nucleotide terminator to the SNP site (Figure 1). This technology provides a simple multiplexable SNP genotyping solution with high accuracy and reproducibility, which is applicable to multiple platforms including capillary electrophoresis, mass spectroscopy, flow cytometry, HPLC, microarray, etc. The CEQ™ 8000 Genetic Analysis System from Beckman Coulter, CEQ DNA Size Standard 80, and CEQ SNP-Primer Extension Kit were newly developed to provide CEQ users an accurate, simple, cheap, and robust solution for SNP scoring and validation based on single-base primer extension technology.

Factor V R506Q (Leiden) Mutation

Screening for genetic variants that predispose individuals or their offspring to disease may be performed at the general population level. Factor V R506Q (Leiden), causing activated protein C (APC) resistance, was discovered in 1994 and is the most common genetic risk factor for venous thrombosis. It is present in 5% of Caucasian Americans, 20% of idiopathic first venous thrombosis cases, and 60% of venous thrombosis cases in pregnant women. In this application note, we demonstrate that the single-base primer extension and CEQ 8000 system provided a novel way to screen this factor V mutation in the human population with a high degree of

Now sold through SCIEX Separations
www.sciex.com/ce

SCIEX

BECKMAN COULTER
Capillary Electrophoresis

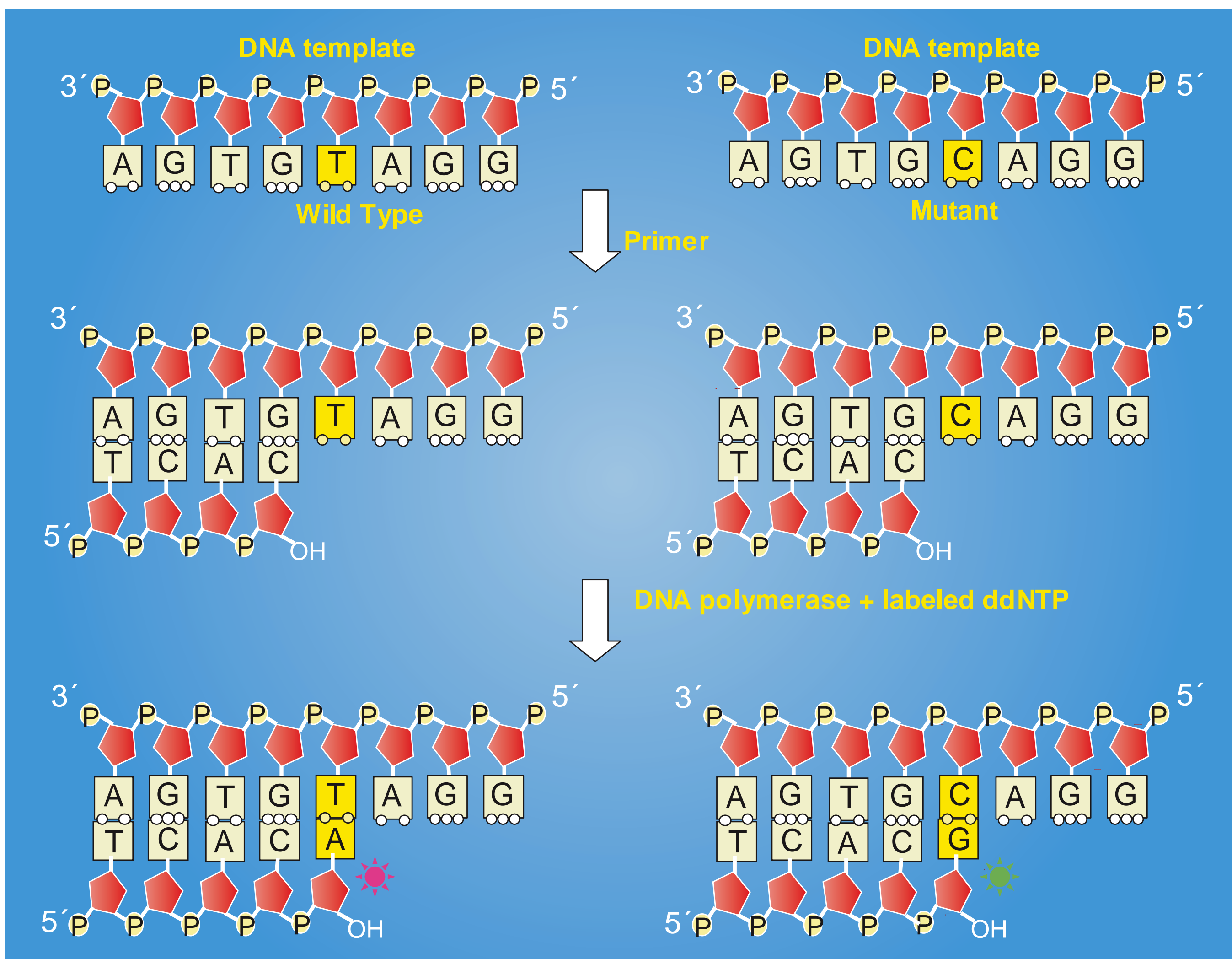


Figure 1. *Primer extension theory.*

throughput, automation, and accuracy. The same method could be designed and applied to validate or score thousands of other mutations associated with human disease in research labs.

Methods and Results

CEQ™ DNA Size Standard 80 (Beckman Coulter P/N 608397)

The CEQ Size Standard 80 was developed specifically for analysis of small fragments on fluorescent detection systems. The CEQ DNA Size Standard Kit-80 contains two reference fragments labeled with the D1 (red) dye for the use in sizing SNP fragments in the approximate size range 20 to 80 nts. It is designed to accommodate a wide range of sizes for up to twelve multiplex or poolplex SNP fragments, and, therefore, provides enhanced flexibility in choosing the length of SNP primers. We recommend the customer to vortex size standard 80 well to achieve equal injection of the two fragments before loading on the CEQ.

CEQ SNP–Primer Extension Kit (Beckman Coulter P/N 390280)

As shown in Figure 2, the primer extension assay contains three major steps: template cleanup, single-base primer extension, and extended product cleanup.

1) Template cleanup: Both plasmids and PCR* products may serve as templates for primer extension. Because PCR products normally contain residual primers and free dNTP, which would lead to generation of non-specific products, it is highly recommended to clean up PCR templates prior to primer extension. The simplest method is to digest the single-strand primers and free dNTP by incubating with Shrimp Alkaline Phosphatase (SAP) and Exonuclease I (Exo I) at 37°C for 1 hour and then inactivate the enzymes at 75°C for 15 min. Both of the enzymes may be added directly into the PCR products without any dilution in a ratio of 6 µL PCR product:2 U SAP: 1 U Exo I.

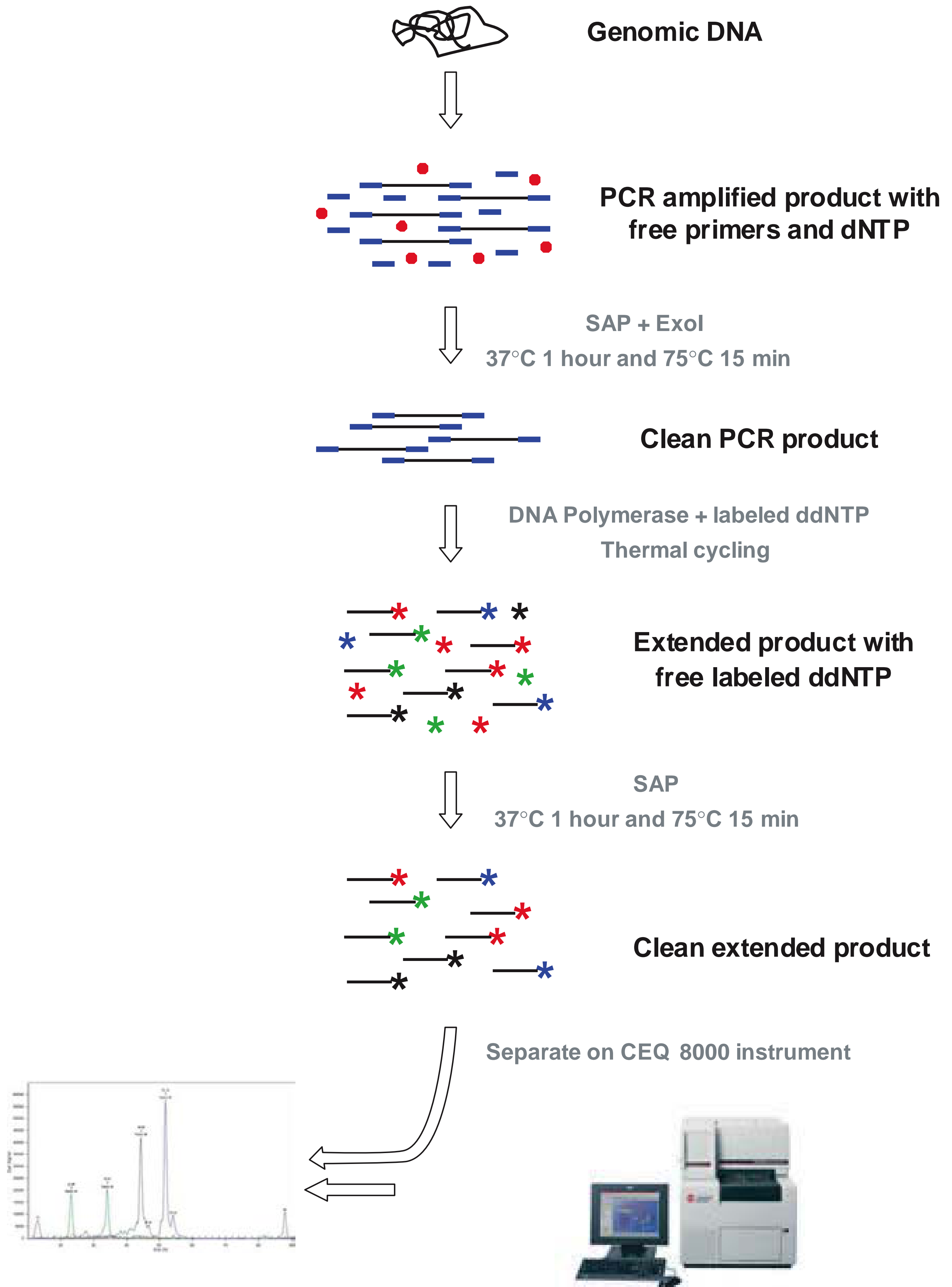


Figure 2. *Overview of primer extension assay and data collection by CEQ™ 8000 Genetic Analysis System.*

2) Primer extension reaction*: The single-base primer extension utilizes the annealing of a regular synthesized DNA primer one base short of the SNP site to DNA template. In the presence of dye-labeled ddNTP, DNA polymerase will add the complementary dye terminator to the 3' end of DNA primer corresponding to SNP site. The primer annealing and extension reactions are normally thermocycled to achieve high signal levels from the extended products. The protocol for setting up primer extension reactions with the CEQ™ SNP–Primer Extension Kit is shown below in Table 2. For multiplex reactions, reduce water accordingly to accommodate all primers and templates. The thermal cycling conditions are shown in Table 3. The annealing temperature is adjustable to 5 degrees below the lowest primer T_m.

Table 1. Prepare SNP-Primer Extension Premix Using CEQ SNP-Primer Extension Kit

Component	*Volume (µL)*
10X Reaction Buffer	210
ddUTP Dye Terminator	210
ddGTP Dye Terminator	210
ddCTP Dye Terminator	210
ddATP Dye Terminator	210
Polymerase Enzyme	105
Total	**1155**

Table 2. Primer Extension Protocol Using CEQ SNP-Primer Extension Kit

Temperature	*Volume (µL)*
SNP-Primer Extension Premix	11.0
SNP primer(s) 1.0 µM	1.0
Template(s) 100 nM	1.0
Water	7.0
Total	**20.0**

Table 3. Thermal Cycling Conditions

Temperature	*Time @ Temperature*	*Cycles*
96°C 50°C 72°C	10 sec 5 sec 30 sec	25
Hold at 4°C	Forever	1

3) Cleanup extended products: The free dye terminators frequently co-migrate with extended product during capillary electrophoresis, which leads to the generation of nonspecific signals. To avoid this interference, each extended product needs to be incubated with 1 unit of SAP at 37°C for 1 hour and then 75°C for 15 min. to remove the 5' phosphoryl groups of the free-labeled ddNTP prior to loading on the CEQ.

We recommend loading 0.5 µL of the reaction samples with SLS and size standard as a starting point (Table 4). If signal over-ranging is observed during analysis, it may be necessary to dilute the samples (1:40) with SLS prior to adding to the above mixture. Cover each sample with one drop of light mineral oil and run the plate on the CEQ 8000 using the SNP-1 separation method (Table 5).

Table 4. Loading Mix

Reagent	*µL/Reaction*
SLS	39.0
Size Standard 80	0.5
Extension Reaction Samples	0.5

Table 5. Run Method: SNP-1

		SNP-1
Capillary	Temperature	50°C
	Wait for Temperature	Yes
Denature	Temperature	90°C
	Duration	60 seconds
Pause	Duration	0 seconds
Injection	Voltage	2.0 kV
	Duration	30 seconds
Separate	Stage 1	
	Primary Voltage	6.0 kV
	Ramp Duration	1.0 minute
	Stage 2	
	Separation Voltage	6.0 kV
	Start Time	1.0 minute
	Ramp Duration	0.0 minute
	Total Separation Duration	16.0 minutes

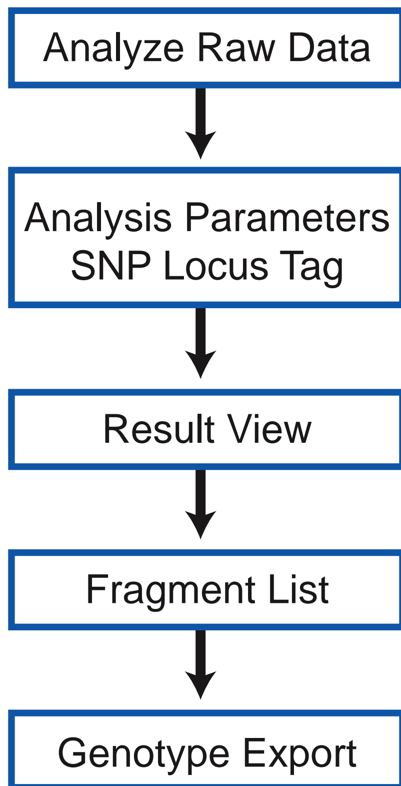


Figure 3. *SNP analysis flow chart using CEQ 8000 software.*

CEQ™ 8000 Genetic Analysis System (Beckman Coulter P/N 285501)

The CEQ 8000 Genetic Analysis System enables the capillary electrophoresis system to perform automated sizing and allele calling from single or multiplexed SNP products. The new size standard 80 provides approximate sizing for SNP locus identification. A new fragment-sizing algorithm was developed to handle multiplexed SNP products. The CEQ 8000 software utilizes a mechanism similar to that for STR loci to analyze SNP. An overview of performing SNP analysis by CEQ 8000 software is shown in Figure 3. The raw data files are analyzed under editable analysis parameters developed specifically for SNP analysis, including SNP Locus Tags, SNP Dye Mobility Correction, Size Standard-80, etc. (*CEQ 8000 User Manual,* P/N 608315). Background peaks may be excluded by modifying the slope threshold. The analyzed results display loci and alleles in the fragment list. No more than one fragment labeled with a given dye may belong to the same locus, while up to four peaks may be labeled as alleles of the same locus, but each must be labeled with a different dye. Dye mobility corrections are applied during data analysis. Multiplexing may be achieved by spacing between adjacent SNP products by 4 to 8 nts.

PCR primer (up)

CCTATACCTTATAAGTGAACATCTTAGAGTTTGATGA**ACCCACAGAAAATGA**
TGCCCAGTGCCTTAACAAGACCATACTACAGTGACGTGGACATCATGAGAGA
CATCGCCTCTG GGCTAATAGGACTACTTCTAATCTGTAAGAGCAGAATCCC

P38 Factor V: C

P28 Factor V: T

Factor V Control

A A

TGGACAGGC G AGG A ATACAGGTATTTGTCCTTGAAGTATGGAAAGTCTTT

AAGACTCTTAAAGAAGACGTTAGG**TCTCCTGGCTAAATAATGGGGCA**TTT

A :P46 Factor V

PCR Primer (down)

G :P56 Factor V

Figure 4. *The factor V mutation locus.*

The software identifies and stores SNP locus information in SNP locus tags stored in the database. The SNP locus tags label the SNP fragment peaks with the correct size and the correct dye. SNP genotypes are summarized and reported in the fragment list through automated SNP locus tag assignment with a high degree of accuracy.

Factor V Mutation Scoring by CEQ™ 8000 Software

As shown in Figure 4, three primers (27 mer, 37 mer, and 47 mer) were synthesized to detect the allele identity at the factor V mutation site. A fourth primer (55 mer) was designed to serve as a positive control. True multiplex primer extension was performed using all four primers. The reactions added one nucleotide to each primer and resulted in products that were 28, 38, 48, or 56 nucleotides long, respectively. Genotypes of DNA template, wild-type, heterozygous, or homozygous for this G to A substitution locus, were scored singly or by multiplex primer extension assays and CEQ 8000 SNP analysis. Sample data were collected and analyzed using default SNP analysis parameters which were developed specifically for analyzing fragments smaller than 100 nucleotides. The apparent sizes for each extended fragment with actual size of 28, 38, 48, or 56 nucleotides were entered to create new locus tags. All sample data were then reanalyzed with four locus tags selected. As shown in Figure 5, the CEQ 8000 software was able to assign the correct allele ID and locus name to the correct peak. For heterozygous loci, the two dye-labeled fragments were labeled by two allele IDs with the same locus name. Over 500 SNP reactions were performed to score this factor V mutation. The CEQ 8000 was able to call the correct genotypes with a high degree of accuracy (almost 100%). The analyzed results display loci and alleles in the fragment list. Column selector and filter sets are available to edit the fragment list display (see A-1929A, "Strategies for Automating the Review of Data"). Genotypes for each result at each locus are summarized in the fragment list and could be directly printed, copied, or exported as a *.CSV file into Microsoft Excel* (Figure 6).

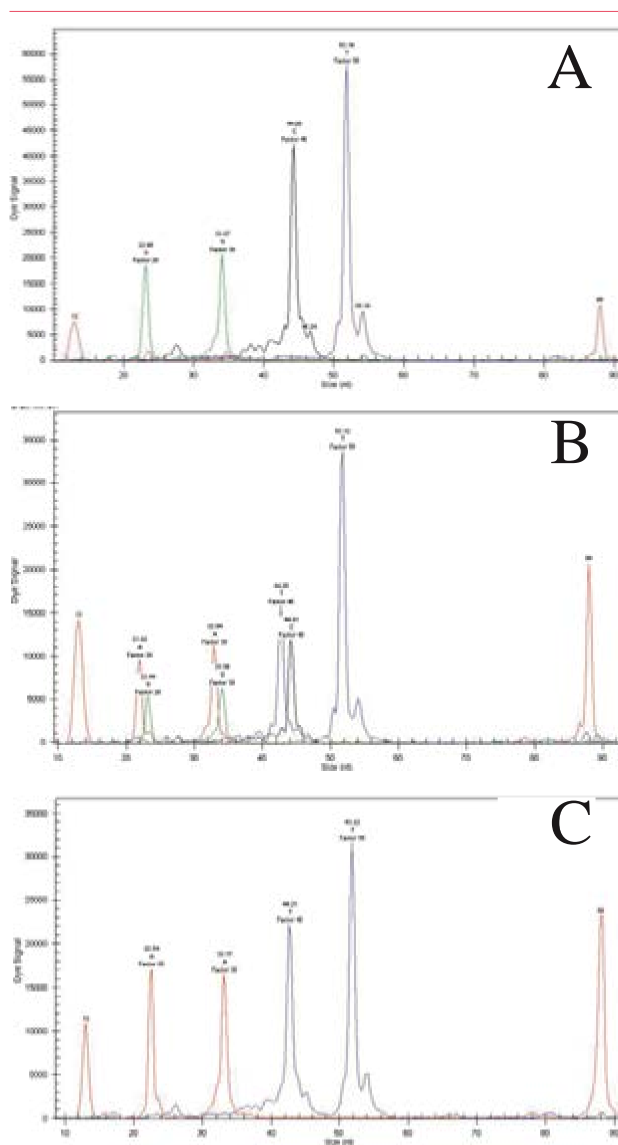


Figure 5. *Correct allele calls for factor V mutation by CEQ 8000 software. A: Wild-type template; B: Heterozygous template at 28, 38, and 48 sites; C. Homozygous template at 28, 38, and 48 sites.*

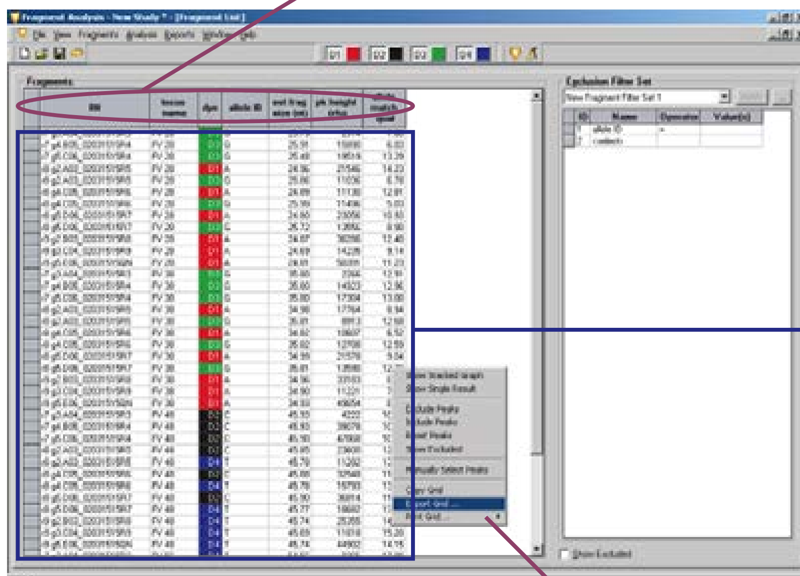
Discussion

To achieve the best results with minimal chemistry optimization, several issues need to be considered:

- The migration of a short fragment during capillary separation is closely related to the length, sequence, and dye of this fragment. Both incorporation of dye and nucleotide composition greatly affect the mobility of the extension products. Often shorter fragments will appear to be nearly five bases longer than their actual size. We strongly recommend short primers be tested before being multiplexed to ensure enough adjacent space for data analysis. A difference of 3–5 nts between primer lengths is recommended.

Utilize Column Selectors to select columns necessary for looking at SNP genotypes

A



Standard SNP genotype summary includes:
- Result name
- Locus name
- Dye
- Allele ID
- Estimated size
- Peak height
- Allele matching quality

Export the fragment list grid as *.CSV file for SNP genotyping summary report

Microsoft Excel - New Study FragmentList.csv

B

	A	B	C	D	E	F	G
1	RN	locus name	dye	allele ID	est frag size (nt)	pk height (rfu)	allele match qual
2	r7 g3.A04_02031515R3	FV 28	D3	G	25.79	2314	7.8
3	r7 g4.B05_02031515R4	FV 28	D3	G	25.91	15890	6.03
4	r7 g5.C06_02031515R4	FV 28	D3	G	25.48	19519	13.29
5	r8 g2.A03_02031515R5	FV 28	D1	A	24.96	21546	14.23
6	r8 g2.A03_02031515R5	FV 28	D3	G	25.86	11036	6.78
7	r8 g4.C05_02031515R6	FV 28	D1	A	24.89	11130	12.81
8	r8 g4.C05_02031515R6	FV 28	D3	G	25.99	11496	5.03
9	r8 g5.D06_02031515R7	FV 28	D1	A	24.8	23056	10.93
10	r8 g5.D06_02031515R7	FV 28	D3	G	25.72	13556	8.9
11	r9 g2.B03_02031515R8	FV 28	D1	A	24.87	36286	12.4
12	r9 g3.C04_02031515R9	FV 28	D1	A	24.69	14235	9.14
13	r9 g5.E06_02031515QN	FV 28	D1	A	24.81	50391	11.23
14	r7 g3.A04_02031515R3	FV 38	D3	G	35.8	2266	12.91
15	r7 g4.B05_02031515R4	FV 38	D3	G	35.8	14923	12.96
16	r7 g5.C06_02031515R4	FV 38	D3	G	35.8	17304	13
17	r8 g2.A03_02031515R5	FV 38	D1	A	34.98	17764	8.94
18	r8 g2.A03_02031515R5	FV 38	D3	G	35.81	8913	12.68
19	r8 g4.C05_02031515R6	FV 38	D1	A	34.82	10607	6.52
20	r8 g4.C05_02031515R6	FV 38	D3	G	35.82	12708	12.59
21	r8 g5.D06_02031515R7	FV 38	D1	A	34.99	21578	9.04
22	r8 g5.D06_02031515R7	FV 38	D3	G	35.81	13590	12.79
23	r9 g2.B03_02031515R8	FV 38	D1	A	34.96	33183	8.56
24	r9 g3.C04_02031515R9	FV 38	D1	A	34.9	11221	7.64
25	r9 g5.E06_02031515QN	FV 38	D1	A	34.93	49654	8.14
26	r7 g3.A04_02031515R3	FV 48	D2	C	45.93	4222	10.52
27	r7 g4.B05_02031515R4	FV 48	D2	C	45.93	39078	10.53
28	r7 g5.C06_02031515R4	FV 48	D2	C	45.9	47868	10.97
29	r8 g2.A03_02031515R5	FV 48	D2	C	45.85	23600	12.02
30	r8 g2.A03_02031515R5	FV 48	D4	T	45.78	11292	13.37
31	r8 g4.C05_02031515R6	FV 48	D2	C	45.88	32948	11.36
32	r8 g4.C05_02031515R6	FV 48	D4	T	45.78	15793	13.43
33	r8 g5.D06_02031515R7	FV 48	D2	C	45.9	36814	11.1
34	r8 g5.D06_02031515R7	FV 48	D4	T	45.77	18682	13.58

New Study FragmentList

Figure 6. SNP genotype summary report. A: Apply filter sets and edit columns to summary SNP genotypes in fragment list. B: Exported genotype report (.CSV) in Microsoft Excel.*

- If signal over-ranging is observed during analysis, it may be necessary to dilute the samples with SLS prior to adding to the above mixture. Different scales of reactions with less enzyme, buffer, and dye terminators are also applicable to reduce signal and save reagents.
- Formation of hairpin or dimmer structures among primers would lead to nonspecific signals after primer extension. To avoid this, we recommend performing a primer-only reaction prior to multiplexing.
- Incomplete removal of the free primer, dNTP, or ddNTP by SAP or Exo I digestion could result in nonspecific fragments. This problem can be avoided by using freshly mixed enzymes for each cleanup.

Conclusions

The CEQ™ 8000 Genetic Analysis System, CEQ DNA Size Standard 80, and CEQ SNP-Primer Extension Kit enable medium- to high-throughput SNP validation and scoring.

- The SNP size standard, SNP dye mobility correction, and SNP locus tag were developed to ensure automated sizing and allele calls for SNP products.
- The CEQ 8000 Genetic Analysis System allows the analysis of multiple SNP loci in a single capillary.
- The CEQ 8000 Genetic Analysis System was able to validate and score human factor V mutation for venous thrombosis population with a high degree of accuracy.

* *The PCR process is covered by patents owned by Roche Molecular Systems, Inc., and F. Hoffmann-La Roche, Ltd. The CEQ SNP-Primer Extension Kit is licensed under U.S. patents 5,888,819 and 6,004,744 owned by Orchid Biosciences, Inc. Excel is a registered trademark of Microsoft Corporation. All other trademarks are the property of their respective owners.*

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.