SoftGenetics Application Note

Automated TILLING® Analysis of Fluorescent Electrophoresis Data with GeneMarker®

Kevin LeVan, David Hulce, Ni Shouyong, Wan Ning, CS Jonathan Liu

Introduction

With the 2007 Science breakthrough of the year going to Human Genetic Variation (1), the importance of methods for studying variations are at an all-time high. The techniques of Targeted Induced Local Lesions In Genomes, or TILLING, and EcoTILLING have been widely used since 2000 to detect Single Nucleotide Polymorphisms (SNPs) in organisms ranging from Arabidopsis (2) to zebrafish (3). The test samples may be experimentally mutagenized (ethylmethanesulfonate, radiation, etc) or from natural populations or derived from tumors or diseased tissues.

Briefly, the genes of interest are identified with gene-specific primers and PCR amplified. The amplicon's primers are labeled with two fluorescent dyes – the forward is often labeled with FAM-blue and the reverse primer is often labeled with HEX-green. The samples are mixed so heteroduplexes can be formed. The hybridized fragments are cleaved at the heteroduplex site by CEL I or Surveyor™ Nuclease, generating multiple pairs of fragments of complementary length and dye color (4). The denatured samples can be run through gel electrophoresis or mixed with an internal size standard and run through capillary electrophoresis. SNPs will yield two peaks of different color and the sum of the sizes will equal the amplicon length (see Figure 1).
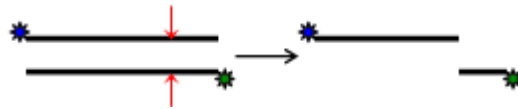

Figure 1: Double-stranded DNA with different colored dyes at each 5′ end will be cleaved at the heteroduplex indicated by the red arrows, yielding two fragments of different color and having sizes that total the amplicon length.

Visual inspection, with the aid of GelBuddy© software, is a common method for determining all locations of point mutations that have been cleaved at the heteroduplex site (5). This leaves some ambiguity due to the efficiency of cleavage and level of noise. GeneMarker contains an application that has been optimized for detecting SNPs generated by the TILLING technique. By using the internal size standard to align each capillary and generating a reference trace from all of the samples in the run, the reference can be subtracted from each individual sample trace yielding a plot highlighting the SNPs. Additionally, a table shows what the expected size of the complementary fragment should be to determine if each peak is a true variation.

Procedure:

1. Launch GeneMarker software.
2. Click Open Files Icon, Click Add and select data (up to 1000 lanes maybe added)
3. Click open and then OK
4. Select Tilling Analysis from Applications menu
5. Select Size Standard, enter Data, Process options and select data size range including whole fragment of amplicon.
6. Click OK to process data

Figure 2: Tilling Options
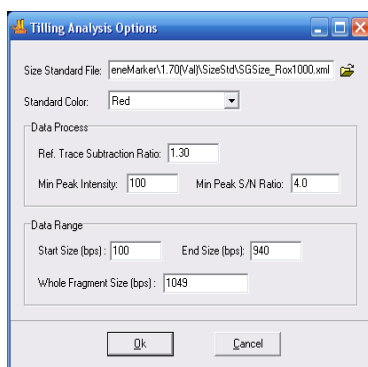
Note: When using gel electrophoresis with a LI-COR or similar platform, you can use SoftGenetics' JelMarker software to create sample files that have a synthetic internal lane standard for each well. GeneMarker can also read the output formats of other instruments, including FSA, AB1, ABI, SCF, RSD, ESD, SMD, and SMR.

Results:

When using GeneMarker's Tilling Application module, the size standard peaks are identified and all sample peaks are aligned. The peaks are smoothed, baseline is subtracted, and lane intensities are normalized. Low quality data is automatically rejected. A synthetic reference trace (Synthetic Control Sample) is constructed using median peak intensities from all of the high quality traces. This reference is subtracted from each sample trace generating a Mutation Chart that automatically identifies the sample's variations as shown in Figure 3, with the amplicon length of 1049bps.
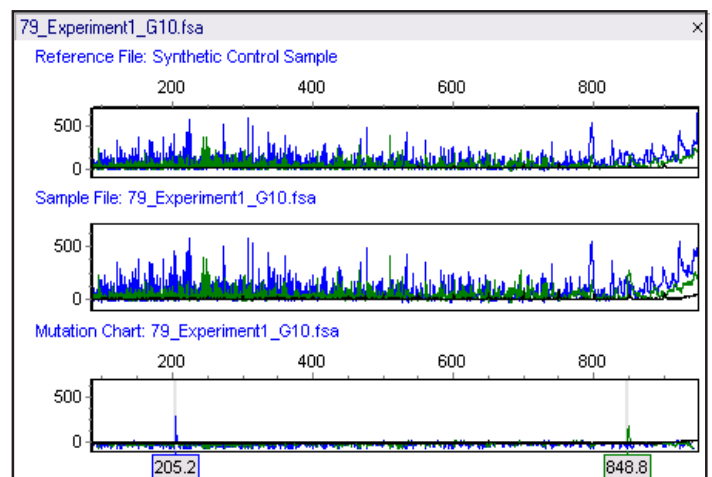

Figure 3: The top panel shows the Synthetic Control Sample obtained from the median intensity after peak alignment. The middle panel shows sample 79. The bottom panel shows the Mutation Chart, generated by subtracting the reference from the sample, identifying individual variations. A blue peak at 205.2bps and a green peak at 848.8bps have been automatically identified. The original amplicon size is 1049bps.

In addition to displaying synthetic gel image, reference trace, sample data file and mutation chart file, the software generates a peak table for active sample and a mutation report for the entire project. Both the peak table and report utilize the assigned amplicon size to predict each peak's complementary fragment size, as shown in Figure 4, simplifying the identification of false positives. The report table, which contains adjustable settings, can be saved in text and Excel formats.

Report

No.	Sample	Dye	Size	cSize	Height	Area	S/N Ratio	Comment
14	73_Experiment1_A10.fsa	Blue	285.3	763.7	777	4388	20.94	
15	73_Experiment1_A10.fsa	Green	278.7	770.3	536	10176	15.89	
16	79_Experiment1_G10.fsa	Blue	205.2	843.8	295	1466	10.43	
17	79_Experiment1_G10.fsa	Green	200.2	848.8	190	3093	6.72	

Figure 4: Blue peaks are displayed in the Sizes column and green peaks are displayed in the cSize column. The complement size (with gray font) is predicted and displayed in the other column.

High Specificity

TILLING is an effective technique for SNP detection, but it can be time consuming. In addition, low yield SNPs might be missed by visual inspection. To assist with both of these factors, SoftGenetics' GeneMarker package with the Tilling Application Module can speed up and increase the accuracy of this process. By creating a Synthetic Control Sample from all of the files in the project and subtracting this reference from each individual sample, variations can be observed more quickly. A report of all variants is generated instantaneously, and a Mutation Chart is shown for each sample file, giving a visual representation of each difference. In Figure 5, lane 73 contains an easily visualized SNP while lane 79 contains a SNP that has a low yield and may be missed – GeneMarker takes the guesswork out of the analysis.
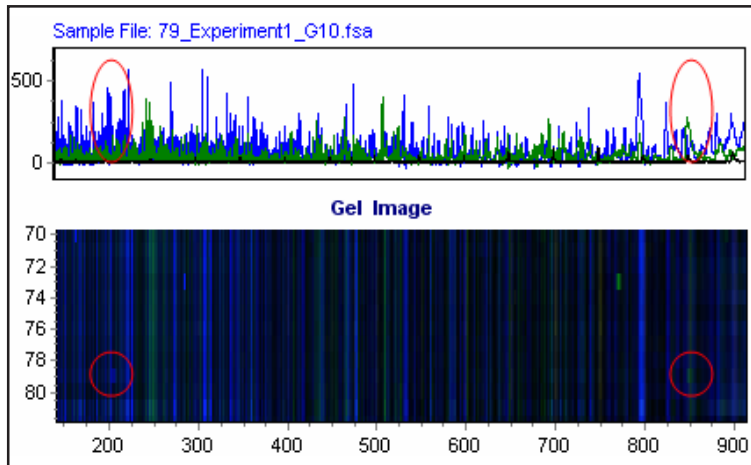

Figure 5: In the Gel Image, lane 73 contains an easily visualized SNP with a blue peak at 285bps and a green peak at 770bps. Lane 79 contains a SNP of much lower yield that could easily be missed during visual inspection.

High Sensitivity

Increased sensitivity is a big advantage to using GeneMarker. EcoTILLING samples, which may contain one mutated portion out of 12 pooled samples, may yield very low levels of the cleaved fragments. We are able to detect variations in EcoTILLING samples where the variant may contribute to only 4% of sample. Some researchers have found 30% more DNA variations by using this tool. In some cases, the intensity of the SNP may fall below the level of noise and making visual inspection impossible – these variants can now be found when the reference noise is removed from the sample. Because the sum of the sizes of each color should equal the whole amplicon size, we are able to detect the cleavage sites in EcoTILLING sample using GeneMarker's Tilling Application module.

Decreased False Positives

False Positives, often caused by chemistry error, random cleavage and instrument spikes, can easily be identified and removed. Spikes often show multiple colors at the same location. The peaks caused by these random processes seldom occur in a complementary size.

Acknowledgement

References

1. **Breakthrough of the year: human genetic variation**. E Pennisi. Science. 2007. **318** (1842-1843).
2. **Targeted screening for induced mutations**. CM McCallum, L Comai, EA Greene, S Henikoff. Nature Biotechnology. 2000. **18** (455-457).
3. **Methods for reverse genetic screening in zebrafish by resequencing and TILLING**. R Sood, MA English, M Jones, J Mullikin, D Wang, M Anderson, D Wu, SC Chandrasekharappa, J Yu, J Zhang, PP Liu. Methods. 2006. **39** (200-227).
4. **A protocol for TILLING and Ecotilling in plants and animals**. BJ Till, T Zerr, L Comai, S Henikoff. Nature Protocols. 2006. **1** (2465-2477).
5. **Automated band mapping in electrophoretic gel images using background information**. T Zerr, S Henikoff. Nucleic Acids Research. 2005. **9** (2806-2812).