

# SoftGenetics Application Note

Feb 2007

## GeneMarker® software for SNPlex™ analysis

David Hulce, Jonathan Liu

### Introduction

Single nucleotide polymorphisms (SNPs) occur frequently and with a relatively even distribution in the human genome (1, 2). Functional SNPs (non-synonymous or nsSNPs occur in the coding region of a gene, regulatory or rSNPs occur in the promoter of a gene) are associated with altered protein function or gene expression (3). Intronic or intergenic SNPs may not alter gene or protein function, but may be associated with complex diseases, drug response or environmental insults (4).

To fully utilize the diagnostic potential of SNPs, a robust high-throughput genotyping and data analysis system should be employed. SNPlex™ Genotyping System (Applied Biosystems) can interrogate 48 SNPs simultaneously and has been used to investigate SNPs in 92 cancer-related genes in breast cancer (5) and to genotype plants (6). GeneMarker genotyping software is designed for fast, accurate and efficient analysis of SNPlex data.

GeneMarker is user-friendly software containing a robust size calling algorithm that detects fragment lengths to less than one base pair and high efficiency allele calling, essential for data generated by SNPlex technology. The software is compatible with electrophoresis systems worldwide, including ABI files, MegaBACE™ and Beckman files, as well as slab gel outputs.

### Procedure

The software settings shown below are recommended to produce the best results for SNPlex analysis. The process from raw data to review of the analysis report occurs in three steps: size calling of the data, importing and adjusting the SNPlex panel, running the sized data against the panel.

Import the data files into GeneMarker using the Open Files icon.

#### Step 1: Suggested Analysis Parameters for Size Calling

Click the Run Project icon to activate the Run Wizard for size calling of the data.

##### Template Selection

1. Panel: none
2. Size Standard: SNPlex\_48plex\_v1
3. Standard Color: orange
4. Analysis Type: SNPlex

##### Data Process-SNPlex Analysis

1. Accept defaults

#### Step 2: Importing SNPlex Panel Information

The panel information can be imported using the Panel Editor (Tools→Panel Editor).

1. Under Files in the menu bar select Import ABI Panels.
2. In the Panel File field navigate to the directory that contains the panel information (file format \*Panels.txt).
3. In the Bins field navigate to the directory that contains the bin information (file format \*Bins.txt).

4. Click OK and the Panel is automatically uploaded to the Panel Editor
5. Rename any of the SNPs, if necessary, to improve readability of final SNP report.
6. To adjust the panel, hold down Shift key and left mouse button, and drag the gray bars over blue or green peaks.

#### Step 3: Compare Sized Fragments to SNPlex Panel

1. Under Panel in the Template Selection window of the Run Wizard use the drop-down menu to select the appropriate SNPlex panel for the analysis (use all previously selected options).
2. To activate the SNPlex analysis report, click Applications in the menu bar and select SNPlex/SNaPshot.

### Results

The SNPlex report window displays a synthetic gel image, list of samples, cluster plot to analyze peak information and assign SNP genotypes, sample electropherogram and SNP genotyping report (Figure 1). The information contained in the report is interlinked. Double-clicking on the cell containing alleles 58/60 for SNP locus P37 of sample A14 highlights the cell, corresponding sample ID in the sample list, data point in the cluster plot and locus in the electropherogram.

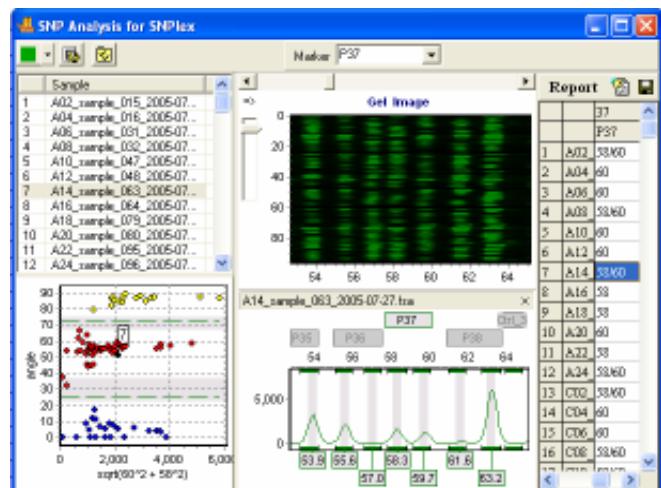
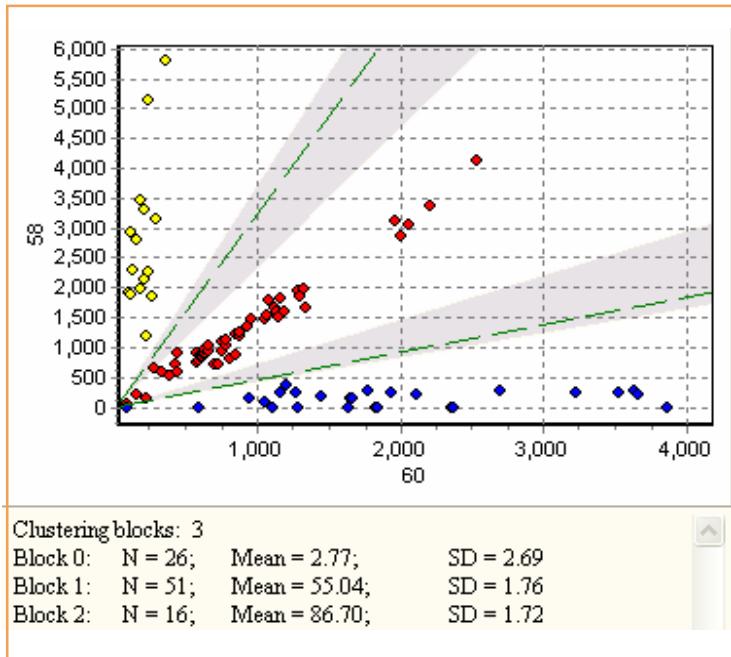


Figure 1: SNPlex report from GeneMarker

The electropherogram displays peak intensity, fragment size (in box along bottom of trace), allele name (along top of trace) and locus identification. Low intensity peaks (<200 rfu) are detected with high sensitivity using derivative to remove influence of baseline. The intensity threshold may be set to filter out low intensity peaks that contribute to incorrect allele calling. A specific SNP is represented by the size and color of the fragments contained in the locus.

The default cluster plot layout is polar view. A Cartesian view (Figure 2) may be selected by clicking the Layout Settings icon on the toolbar. Polar view plots the angle of the vector connecting the peak intensities of the SNP—shorter fragment on y-axis and longer fragment on x-axis—and the origin against the square root of the sum of the square of the longer fragment and the square of the shorter fragment. The number of samples, mean and standard deviation for each group can be displayed under the clustering plot.



**Figure 2:** Cartesian view of cluster plot from GeneMarker

Data points within the gray region of the cluster plot indicate SNP calls of low confidence. The clustering algorithm calculates a t-value for each data point in each cluster. Samples that do not meet the 95% confidence level for the cluster will appear in the gray region. Editing SNPs can be done by double right-clicking on a data point in the cluster plot or on the cell of the report table.

The SNP analysis report can be saved as Excel (\*.xls) or Text (\*.txt) files.

#### Discussion

Various techniques have been developed to interrogate SNPs, including single nucleotide primer extension (SNuPe), SNPlex™, SNPWave™ (KeyGene), SNP chips and DNA sequence analysis. SNPlex has the capacity of high-throughput genotyping.

GeneMarker is highly accurate and easy to use. High efficiency allele calling and low intensity peak detection along with robust size calling and clustering algorithms can quickly and efficiently determine SNP genotypes, maintaining the high-throughput advantages of SNPlex genotyping.

#### Acknowledgments

We would like to thank James Knowles and Oleg V. Evgrafov of Columbia University for providing the SNPlex data.

#### References

- Kruglyak L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics* 22:139-144.
- Venter JC, MD Adams, EW Myers, PW Li, RJ Mural, GG Sutton, et al. 2001. The sequence of the human genome. *Science* 291:1304-1351.
- Salim Mottagui-Tabar, MA Faghihi, Y Mizuno, PG Engström, B Lenhard, WW Wasserman and C Wahlestedt. 2005. Identification of functional SNPs in the 5-prime flanking sequences of human genes. *BMC Genomics* 6:18. <http://www.biomedcentral.com/1471-2164/6/18>.
- SNP Fact Sheet. Human Genome Project Information. U.S. Department of Energy-Office of Science, Office of Biological and Environmental Research. [http://www.ornl.gov/sci/techresources/Human\\_Genome/faq/snps.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml).
- A Vega, A Salas, C Phillips, B Sobrino, B Carracedo, C Ruiz-Ponte, R Rodríguez-López, G Rivas, J Benítez, A Carracedo. 2005. Large-scale single nucleotide polymorphism analysis of candidates for low-penetrance breast cancer genes. *Breast Cancer Research* 7(Suppl 2):P1.14.
- Massimo Pindo, M Troggio, D Cartwright, A Gutin and R Velasco. 2007. SNPlex™ Genotyping System to Develop A Dense SNP-Based Genetic Linkage Map of Grapevine (*Vitis vinifera* L.). Plant & Animal Genomes XV Conference; San Diego, CA. January 13-17:P74.

Trademarks are property of their respective owners.