

**SOFTGENETICS<sup>®</sup>**

**MUTATION**  
**Surveyor<sup>®</sup>**  
DNA Variant Analysis Software

**Fast, Accurate and Sensitive DNA Variant Detection from Sanger Sequence traces:**

- Anti-Correlation Technology Provides 99.5% Accuracy & 5% Sensitivity
- Compatible with all major capillary sequencing platforms and chemistries
- Detects all DNA Variants, SNPs, Indels, Mosaic
- Pipeline Automation- analyze up to 2000 sequence traces in minutes
- Low Frequency Variant Detection and Quantification
- Indel Detection-Deconvolution of Heterozygous Indel
- Mitochondrial Sequence Analysis
- Methylation Analysis
- Customized reporting features-color coding of variants by function

**SOFTGENETICS<sup>®</sup>**  
Software PowerTools for Genetic Analysis  
[www.softgenetics.com](http://www.softgenetics.com)

**Mutation Surveyor®** Software combines both automation and accurate variant calling into a single package. The software is compatible with all major capillary electrophoresis systems, allowing for an integrated workflow from importing of raw data to analysis and reporting. The patented anti-correlation technology physically compares sample sequence traces to a reference trace, calculating dropping factor, overlapping factor, peak height and signal to noise ratio of each individual peak, providing accuracy up to 99.5% with Phred 20 bi-directional sequences with a sensitivity of 5% of the primary peak. Additionally, the software monitors the mobility of sample traces to the reference to detect indels ranging from one bp to 15% of the sample sequence.

Mutation Surveyor can analyze up to 2000 lanes of data in less than 15 minutes, provides an unattended pipeline tool that includes a heterozygous indel deconvolution tool, performs hypervariable sequence analysis, mutation quantification, mitochondrial analysis, low frequency variant detection, and methylation analysis. Mutation Surveyor also offers many customizable reports that include color-coding and saving reports in XML format.

**Anti-Correlation Technology:** Mutation Surveyor uses anti-correlation technology, which compares sample sequences to a reference sequence. The patented algorithm detects variants and calculates a mutation confidence score by using peak intensities, dropping factor (vertical), overlapping factor (horizontal base spacing), and signal to noise ratio of each peak. Using this technology with bi-directional data, accuracy is over 99.5% for Phred 20 sequence traces with sensitivity to 5% of the primary peak or 1 part diseased in 19 parts normal when analyzing cancerous samples. Found variants are displayed as peaks for fast and easy review in our mutation electropherograms as shown in Figure 1-center panes.

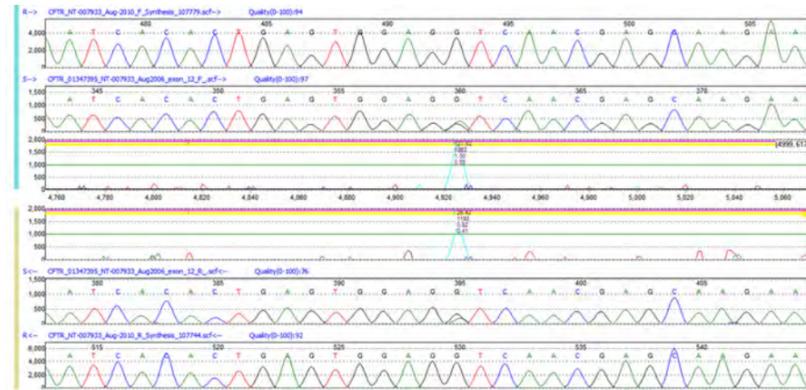
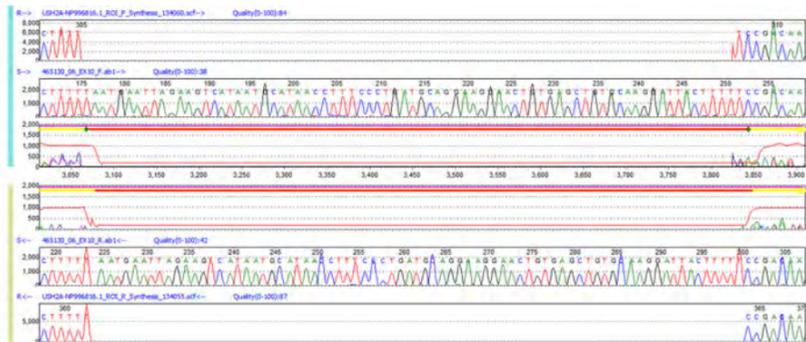


Figure 1

**Multi-Step Sequence Alignment**

SoftGenetics' multistep alignment algorithm allows Mutation Surveyor to detect indels based on mobility shift when the samples are aligned to the reference. The software is capable of aligning sequence traces containing up to 20% variation from the reference. By aligning the traces in 12bp, 6bp, and 3bp-mers to the reference trace, the software can detect both small and large indels, up to 15% of the sequence trace.

Figure 2: Alignment Algorithm Detects Large Indels

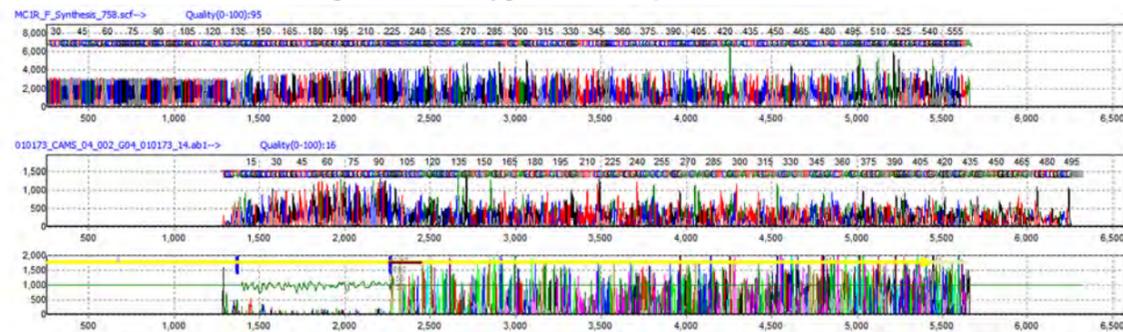


This 76bp insertion was detected in a sample trace that was 500bp long. The insertion accounts for 15% of the sample and is detected by Mutation Surveyor at default settings. Indels are detected by monitoring and comparing the mobility line during alignment. Any shifts in the mobility of the sequence trace could indicate an indel. In Figure 2, the insertion is indicated by a negative mobility line. The software also displays a heavy red line on top of the indel position.

**Heterozygous Indel Deconvolution**

Mutation Surveyor's Heterozygous Indel Deconvolution tool automatically deconvolutes the heterozygous mutant trace from the wild-type trace. Once the indel is detected, it is indicated by a horizontal brown bar in the main analysis window as displayed in Figure 3.

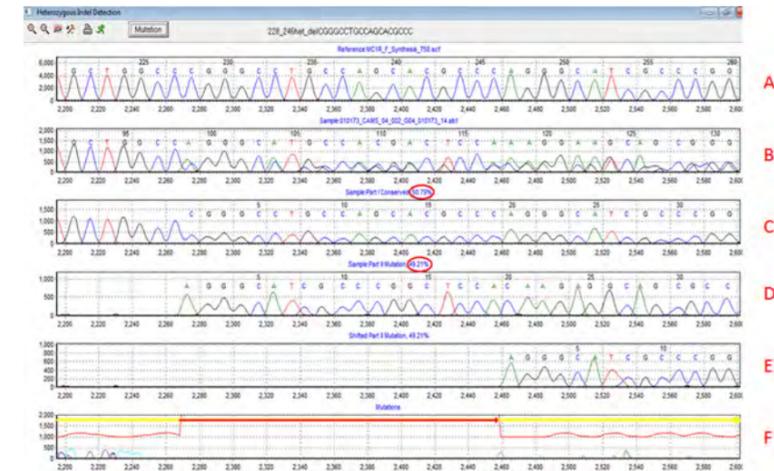
Figure 3: Heterozygous Indel Sequence Pattern



This is a typical pattern for a heterozygous indel. After the indel position, there is a pattern of double traces, one wild-type and one mutant trace with the indel. The Mutation Electropherogram shows that there are less noise within the region of the sample trace that matches well to the reference trace. The noisy region, after the heterozygous indel, is attributed to the overlapping wild-type and mutant traces.

Heterozygous Indel Deconvolution tool separates the mutant trace from the wild-type trace to detect the indel variation and quantification as shown in Figure 4. The entire tool is interactive, variations can be edited, and images can be printed at high resolution.

Figure 4: Heterozygous Deconvolution Tool

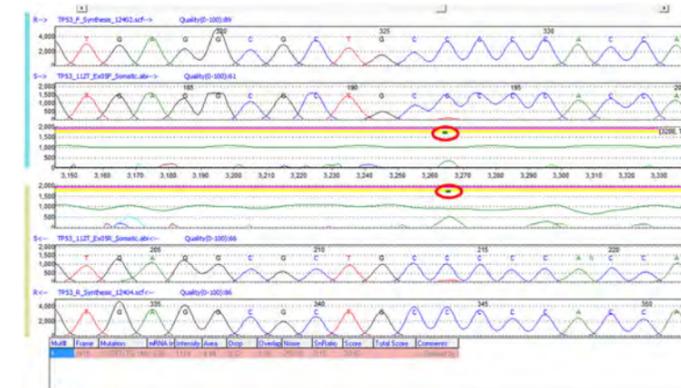


This illustration shows the detection of a 19bp heterozygous deletion. (A) is the reference trace, (B) is the sample trace, (C) is the wild-type trace aligning to the reference, (D) is the deconvoluted mutant trace, (E) is the shifted mutant trace aligning to the reference, and (F) is the mutation electropherogram that indicates the position and type of indel. The mutant percentage is estimated by taking the ratio of the mutant intensity over the total intensity and reported, as indicated by the red circles in Figure 4.

**Low Frequency Variant Detection and Quantification (Somatic Mutation, Heteroplasmy, and Mosaicism)**

Low Frequency variations are often difficult to detect as they can be buried in the baseline of the sequence trace. These mutations include somatic mutations, heteroplasmy (usually in mitochondria) and mosaicism. Mutation Surveyor contains a special algorithm to detect and display these possible low frequency variations.

Figure 5



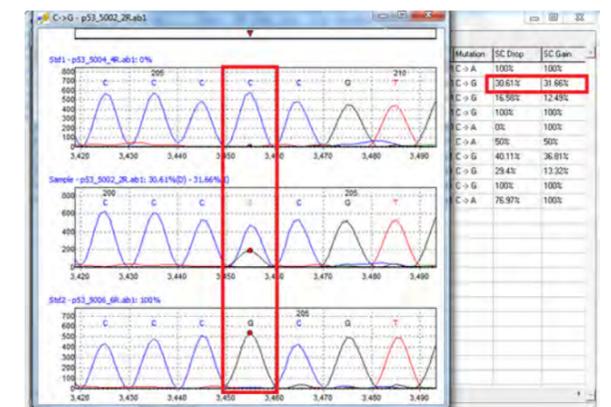
A possible somatic mutation detected in the tp53 gene. The green bar in the mutation electropherogram indicates the position of the possible somatic variation. The software detects these low frequency variations by looking at both the forward and reverse traces. If it detects a minor component of the same wavelength at the same position in both traces, it will alert the user to that position with the green bar.

**Mutation Quantification**

Mutation Surveyor has a built-in Mutation Quantification tool to eliminate the needs of transferring data from one software package to another. This tool offers two methods of quantification, Standardized Allele Ratio and Simplified Allele Ratio.

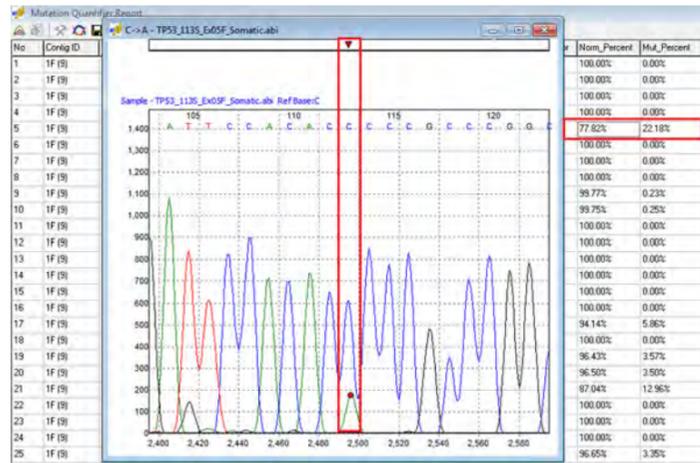
Standardized Allele Ratio will use two standards to quantify variations of interest: mutant and wild-type standards. In Figure 6, the drop intensity of the "C" nucleotide is about 30.61% while the gain in the "G" intensity is about 31.66%. Standard 1, the top pane, is the wild-type sequence. Standard 2, the bottom pane, contains the 100% homozygous mutant sequence. The middle pane contains the test sample.

Figure 6: Standardized Allele Ratio



Simplified Allele Ratio does not require wild-type and mutant standards to calculate the mutant percentage. Instead, it looks directly at the position to estimate the percent contribution of the wild-type allele versus the mutant allele by assuming equal fluorescent intensity of all dye color in a one to one allelic ratio.

Figure 7: Simplified Allele Ratio

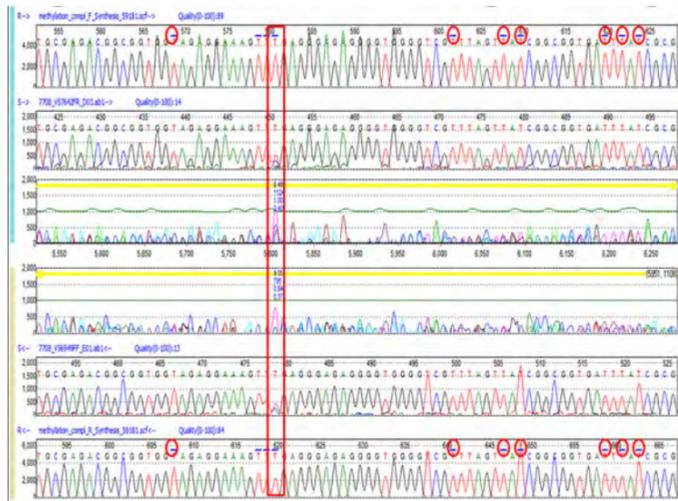


This figure shows the quantification of a possible somatic mutation in the tp53 gene. Mutant percentage is estimated to be at 22.18% while the normal is 77.82%. This quantification tool can be used to calculate processed mutation calls, position of interests, and possible somatic mutations.

### Methylation Analysis

Mutation Surveyor has the option to analyze sequences that have gone through bi-sulfite treatment for methylation studies. After sequencing, the resulting files can be imported straight into Mutation Surveyor for analysis. The software will automatically identify methylated sites, successful, unsuccessful and partial conversion of Cytosine to Uracil sites during bisulfite treatment. It takes one click to activate the methylation analysis parameters.

Figure 8: Methylation Analysis Display



Methylation analysis in Mutation Surveyor software displays the positions that used to be a cytosine with a blue bar on top of the position. The indicated variation shows incomplete conversion of cytosine to uracil at that position.

Mutation Surveyor software has a specialized methylation report that indicates all positions that are methylated, successfully converted, partially converted, and unsuccessful conversion. In Figure 9, the samples are listed at the top row, positions and nucleotides are in the first column. You can double click on any cell in the report and it will take you to the specific position in the main analysis window for quick review. This report can be printed, saved, or exported as a text or excel format file.

		7373_V56949FF_B01.ab1	7373_V57642FR_A03.ab1	7708_V56949FF_E01.ab1	5505_V56949FF_C01.ab1	7708_V57642FR_D03.ab1
58876	G	g	-	-	-	-
58877	G	g	-	-	-	-
58878	G	g	-	-	-	-
58879	G	g	-	-	-	-
58880	T	t	-	-	-	-
58881	C	t	S	S	S	S
58882	C	c	M	M	M	M
58883	G	g	-	-	-	-
58884	C	t	S	S	S	S
58885	C	t	S	S	S	S
58886	T	t	-	-	-	-

Figure 9

### Hypervariable Sequence Analysis

Using Mutation Surveyor's multi-step alignment algorithm, the software can align hypervariable sequences to the correct reference for mutation detection. The graphic display of the main analysis window displays all called mutations, a mutation table, amino acid changes and positions of differences for easy reviewing and interpretation. This application allows users to analyze virus DNA sequences, HLA typing, and many other hypervariable sequences.

Figure 10 shows a HCV sample with high mutation frequency aligning to the reference for mutation detection. Variations in the sample that causes a change in the amino acid can be seen in (1) the amino acid pane. The mutation electropherogram (2) indicates positions of differences between the two traces with a peak at the location. Each mutation call is recorded in the mutation table (3) at the bottom. For further clarification, the GenBank pane (4) also displays the nucleotides differences with an asterisk at the bottom of the nucleotides.

Figure 10: Hypervariable Sequence Analysis

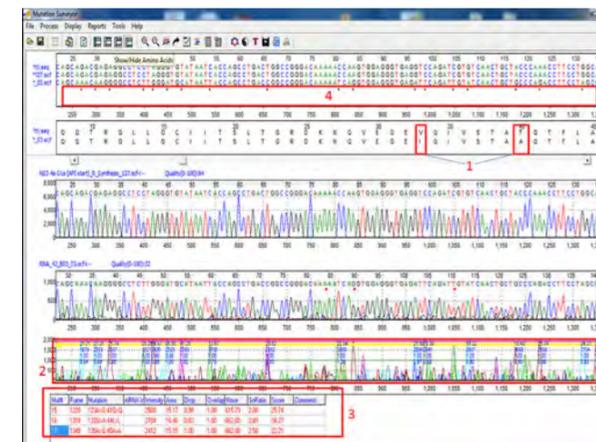


Figure 11

### Reference Assembly

The Project Reviewer Tool can perform reference assembly on replicate samples or overlapping amplicons that align to the same reference. Users can view the entire project with multiple electropherograms and samples in the same window.

Project Reviewer: main analysis window showing overlapping amplicons of a single patient with color-coding of variants. cDNA numbering is shown with the nucleotide string if it is present in the Reference GenBank or SEQ files. Users can customize the display by choosing which panes to use by using the toolbar at the top of the Project Reviewer.

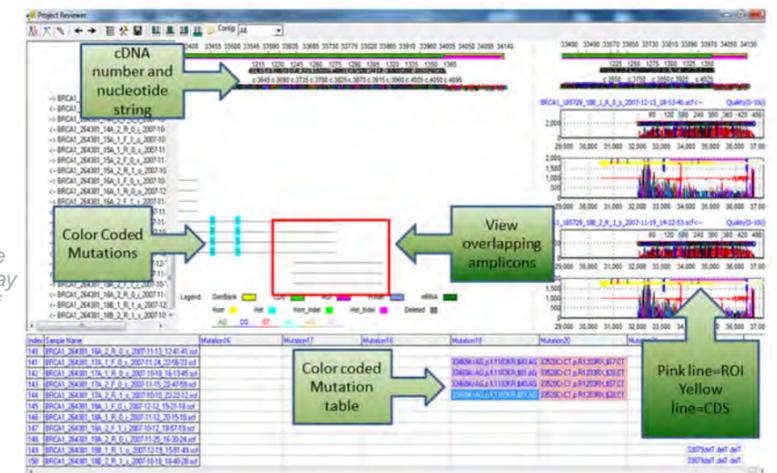
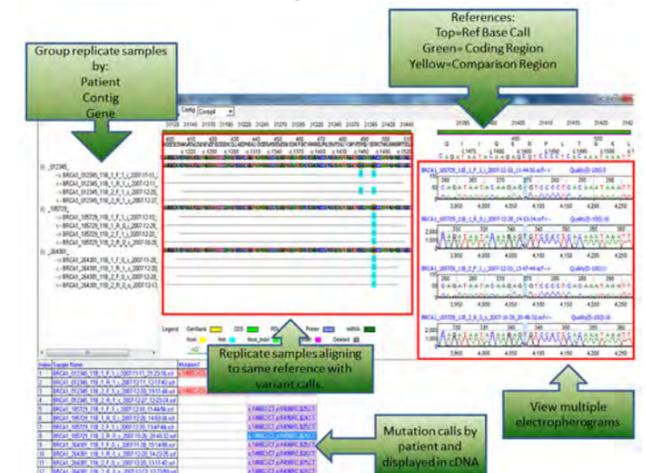


Figure 12

The Project Reviewer also has a grouping tool that allows the grouping of sequences by characters or by amplicon ID. In Figure 12, the illustration shows replicate samples grouped by patient IDs.

Project Reviewer showing replicate samples for three different individuals aligning to the same reference. The samples are grouped by patient ID. Variants are detected at the same positions in all 3 patients. Multiple electropherograms can be seen at one time to compare mutation calls between samples. Mutation table at the bottom lists all mutations for the selected patient.



## Mitochondrial DNA Sequence Analysis

The mitochondrion genetic code has a few key differences from the human genetic code. Mutation Surveyor software will automatically switch to the mitochondrion genetic code if a mitochondrion GBK file is loaded. It will use this genetic code to show the correct amino acid change when analyzing the sample sequences.

Mutation Surveyor software can automatically read a tag in the GenBank file to recognize the mitochondrion genetic code and use it for the analysis without the need for any additional steps. The software can also display multiple CDS, as shown in Figure 13 (indicated by the yellow line), for the amplicon if annotation is available in the GBK file. The start and end position of the mitochondrion's circular genome is indicated by the double sided arrow in Figure 14. The amplicon in the illustration spans the origin of the genome and the sequence can match continuously.



Figure 13

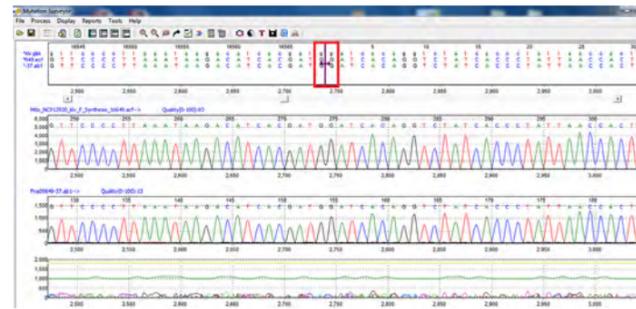


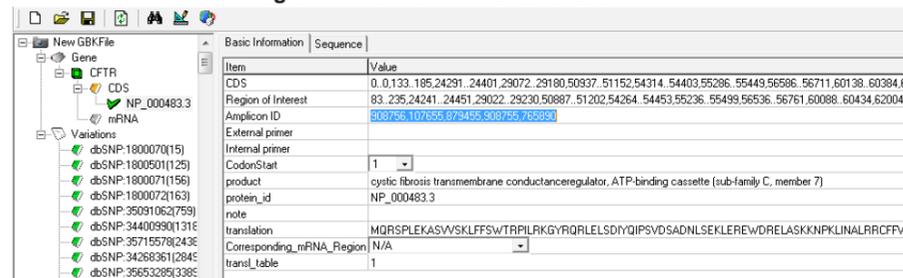
Figure 14

## GenBank and SEQ File Editors

Mutation Surveyor software package also comes with GenBank and SEQ file editors so that Regions of Interest and variants can be added for analysis. These files can also be extracted from projects to be used for future references.

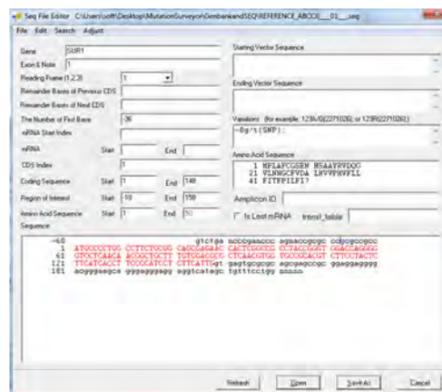
The Advanced GBK File editor can open any GenBank file downloaded from NCBI. In this dialog, the user can choose specific isoforms that will be used for the analysis, create Regions of Interest, and add variants to the variation list.

Figure 15: Advanced GBK File Editor



Advanced GenBank File Editor can add specific Regions of Interest or have the software automatically create ROIs based on CDS. Amplicon IDs can be added in the Amplicon ID field. The variations in the GenBank file can also be edited by adding, deleting, or editing an existing variation for personal needs.

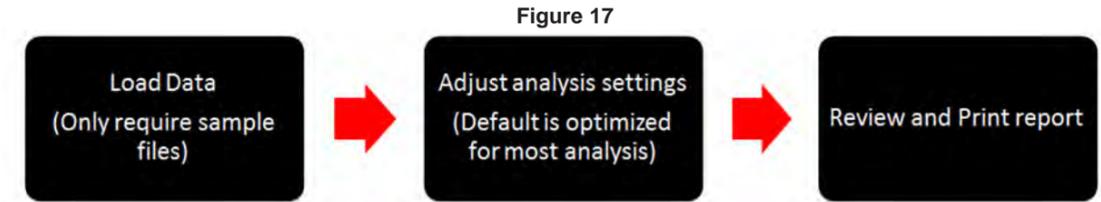
Figure 16: SEQ File Editor



The SEQ File Editor allows the user to import individual SEQ files for editing. In this dialog box, the user can specify the ROI, CDS, Amino Acid, Variations, etc... This tool is used to create individual SEQ files for a single amplicon to be used for mutation detection. After saving, the file can be imported directly into the software.

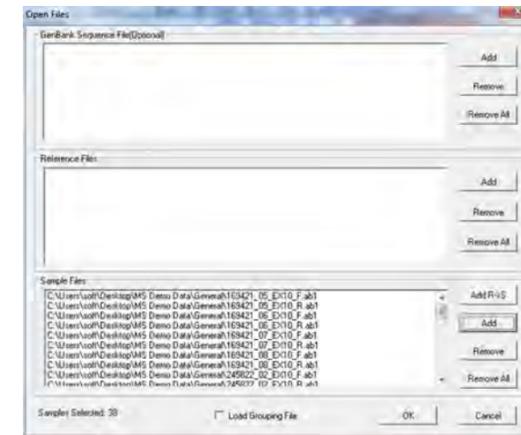
## Automation Pipeline:

Mutation Surveyor software features an Automated Analysis Pipeline.



When importing files, the only files required are sample files as shown in Figure 18. GenBank files, used for annotations, will be automatically downloaded if "GenBank Sequence File" pane is left blank. Reference files, used for mutation detection, are automatically synthesized if the "Reference Files" pane is left blank.

Figure 18



Mutation AutoRun is the tool in Mutation Surveyor that is used to set up batch processing. It only takes 4 steps to set up multiple jobs to be analyzed:

1. Specify each job name
2. Specify GenBank/SEQ files directory (optional)
3. Specify Reference files directory (optional)
4. Specify Sample files directory

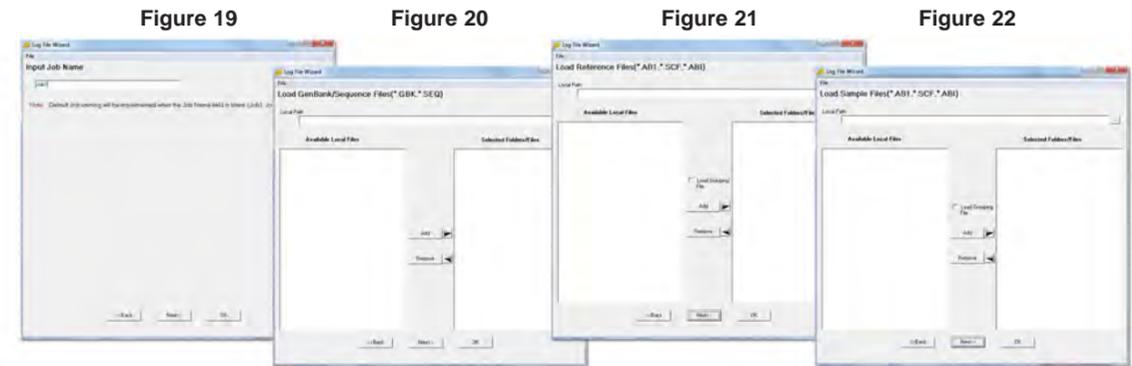
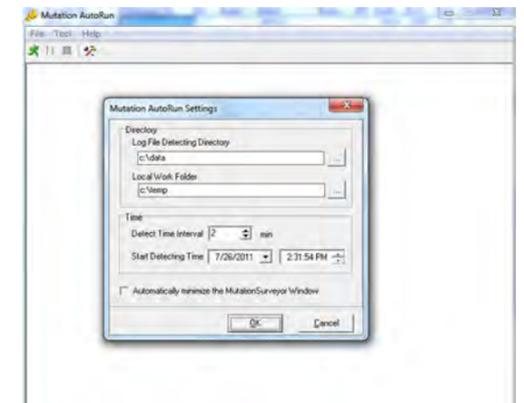


Figure 23

Mutation AutoRun settings allow you to specify the input and output directory of log files and finished project. It also allows you to set the start time and time intervals in which Mutation Surveyor checks the directory for new log files for analysis.



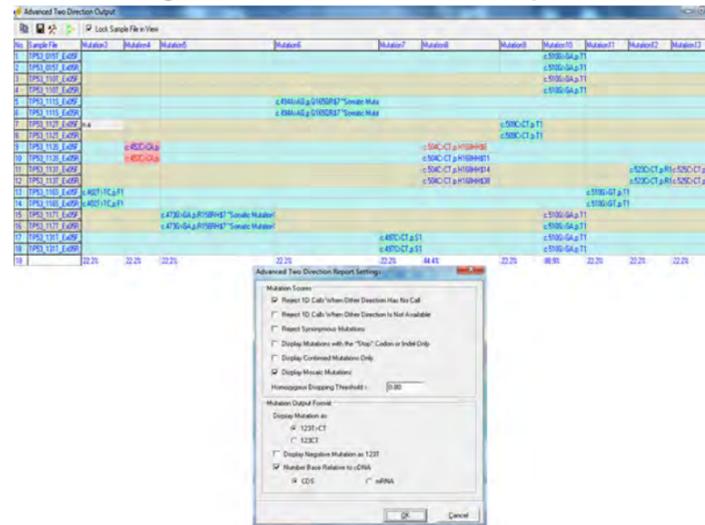
## Reporting Options

Mutation Surveyor offers a variety of reporting options including an entirely customizable report. The types of reports range from unique analysis, such as the Methylation Report, to more generalized reports for 2-directional data. Some of the reports contain color-coding for variant tags, grouping options, and filtering options for different types of variations. Reports are linked to graphics in order to easily view variants of interest.

## Advanced 2-Directional Report

The Advanced 2-Directional Report has options to ignore calls only made in one direction as they tend to have a lower confidence. Within the report, there are different display and filtering options.

Figure 24: Advanced 2-Directional Report



Somatic mutations are displayed in the report with a "Somatic Mutation" tag. Variations can also be displayed based on the cDNA number as offered in the settings. The user can double click on any variation in the report and the main analysis window will zoom into that position.

## Custom Report

The Custom Report is the most versatile report in Mutation Surveyor. It has options to customize everything from grouping of the samples to which type of mutations are displayed, as shown in Figure 25. It also has filtering options, by Coding Sequences (CDS) and Regions of Interest (ROI), for mutation calls. In addition, there are color-coding options, shown in Figure 26 and 27, for variant tags associated with variations detected for easy identification. The report can be printed or exported as a text, excel, or xml format.

Figure 25: Mutation Filter Options

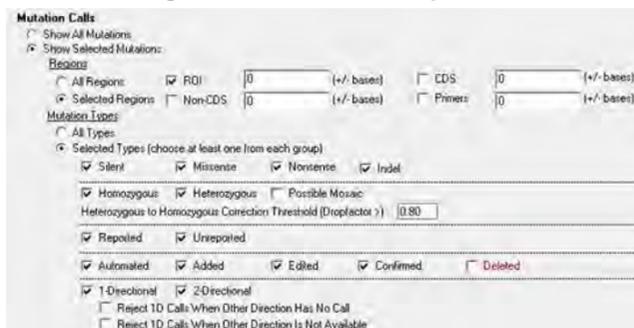
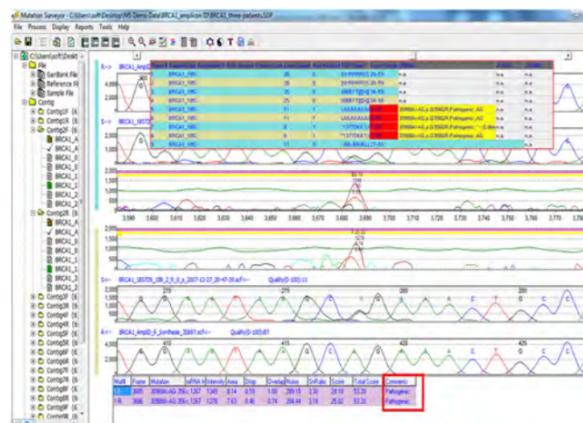


Figure 26: Mutation Filter Options



Figure 27: Custom Report Display

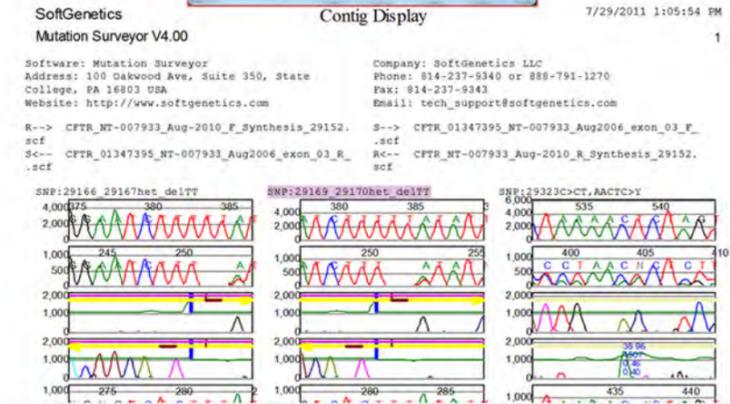
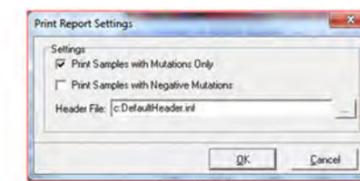
Variant detected in the illustration has a "pathogenic" tag. This specific tag was selected to be color-coded yellow in Figure 27 and shows up in the report with a yellow background.



## Clinical Report

The clinical report has been formatted to only display samples that contain mutation calls or reported variants based on the GenBank file. The report allows for a custom header that may include the institution name, address, and other information.

Figure 28: Clinical Report



Clinical Report is displaying the header information at the top and the mutation calls made in the sample as well as the sample names. Each page of the report refers to a different sample and the mutation calls made in those samples. This report can be printed as high-resolution images.

## Other Reports Include:

- Mutation Report
- Project Report
- HGVS Report
- Graphic Display Report
- GAD Report
- MSM\_Song Report
- Full Gene Report
- PrettyBase Report

Please open disk to review your applications of interest and to install a 30-day trial of Mutation Surveyor software.

## Disk Contains:

- Application Notes
- 30-day free Mutation Surveyor Trial
- User Manual

Mutation Surveyor software has been tested and validated for various computer systems. The minimum system requirements are:

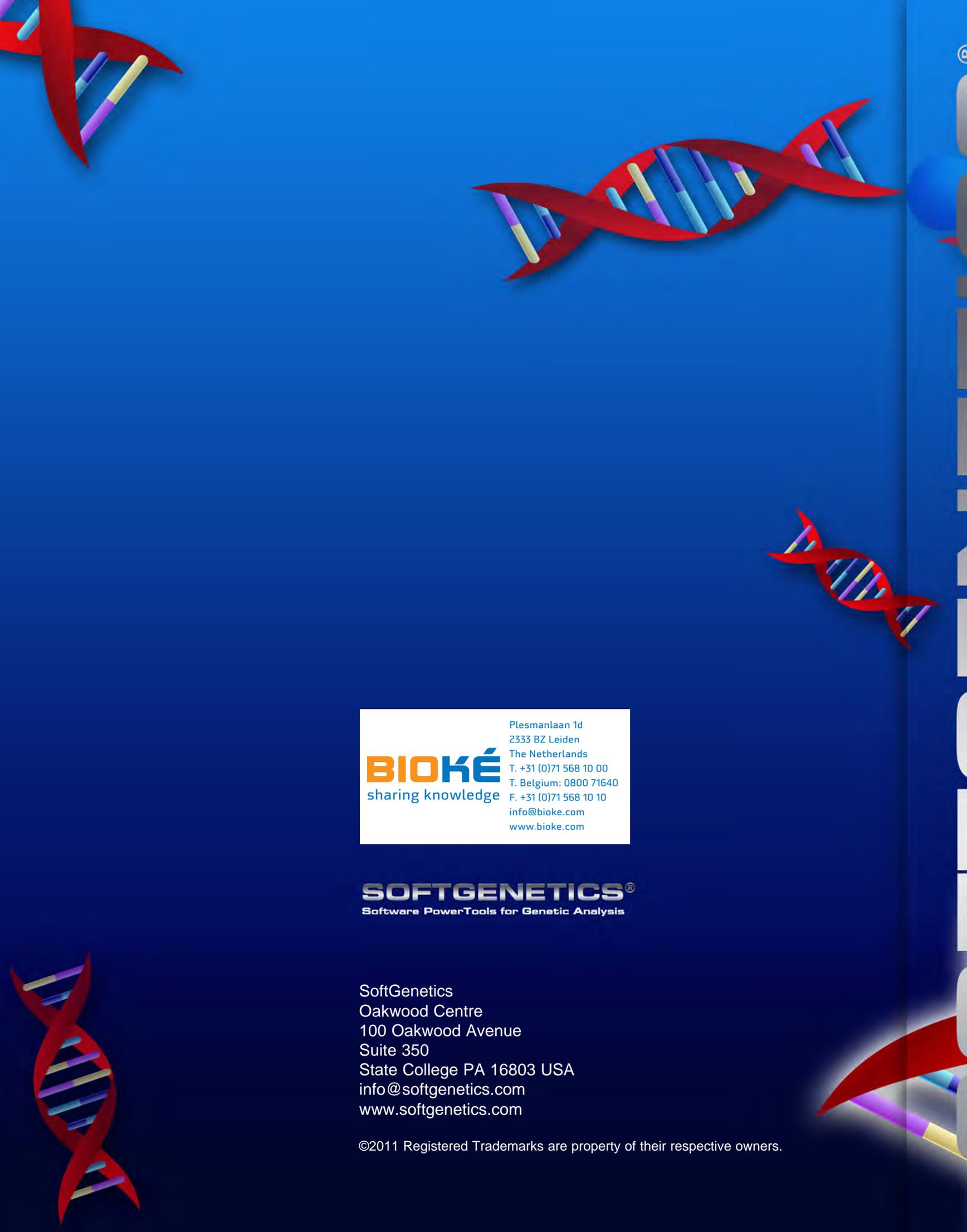
## Windows® PC

- OS: Windows® XP, Vista, Windows® 7
- Processor: dual-core, 1.8GHz
- RAM: 2 GB
- Available hard disk space: 60GB

## Intel® Powered Macintosh®

- Parallels® desktop for Mac (Mac OS/virtual machine dependent) or Apple™ Boot Camp or VMware® Fusion (Mac OS/virtual machine dependent)
- RAM: 2GB
- Available hard disk space: 60GB

Installation of Mutation Surveyor is not supported on Linux or UNIX-based operating systems.



**BIOKÉ**  
sharing knowledge

Plesmanlaan 1d  
2333 BZ Leiden  
The Netherlands  
T. +31 (0)71 568 10 00  
T. Belgium: 0800 71640  
F. +31 (0)71 568 10 10  
info@bioke.com  
www.bioke.com

**SOFTGENETICS®**  
Software PowerTools for Genetic Analysis

SoftGenetics  
Oakwood Centre  
100 Oakwood Avenue  
Suite 350  
State College PA 16803 USA  
info@softgenetics.com  
www.softgenetics.com

©2011 Registered Trademarks are property of their respective owners.