

Detection of Insertions and Deletions, Including Deconvolution of Het-Indels in Sanger Resequencing Projects using Mutation Surveyor® Software

Brent DiGiorgio, Kevin LeVan, ChangSheng Jonathan Liu

Introduction

Insertions and deletions (INDELS), also known as deletion and insertion polymorphisms (DIP) and deletion/insertion variations (DIV), may create a new stop codon and terminate the amino acid translation, resulting in severe genetic diseases. Although Next Generation sequencing techniques have increased in popularity and proven to be useful, Sanger Sequencing remains the gold standard for genome sequencing projects. These large resequencing projects demand fully automated calls of the heterozygous indels in determining the disease causing mutations. Among the diseases caused by indels are Tay-Sachs disease, Alzheimer's disease¹, and various forms of cancer^{2,3}.

Mutation Surveyor is a unique software package which detects both homozygous and heterozygous INDELS. To detect INDELS Mutation Surveyor uses a proprietary algorithm which compares the mobility of the sample trace to the reference trace, and when a change is noted gaps the reference or sample trace until realignment is possible. This technology has proven to detect INDELS with greater than 99% accuracy⁴ when bi-directional sequences of greater than Phred 20 are employed.

Additionally, when a heterozygous INDEL is detected the software will automatically deconvolute the mixed trace into two clear traces, eliminating the need of tedious and error prone manual deconvolution. Recent studies by Ellard et al.⁵ illustrate a sensitivity >99% when using Mutation Surveyor for determining heterozygous base substitutions.

Mutation Surveyor software features an easy to use interface, a low learning curve, full automation, and an exclusive migration time comparison to detect indels. Mutation Surveyor's patented physical trace comparison (U.S. Patent No. 7,400,980) is able to regularly and accurately detect various sizes of indels, with studies showing detection of 3 base pair indels⁶ to 30 base pair indels.^{7,8} The program also offers a unique Heterozygous Indel Detection tool, with sensitivity at 20% contribution from the mutant/novel allele.

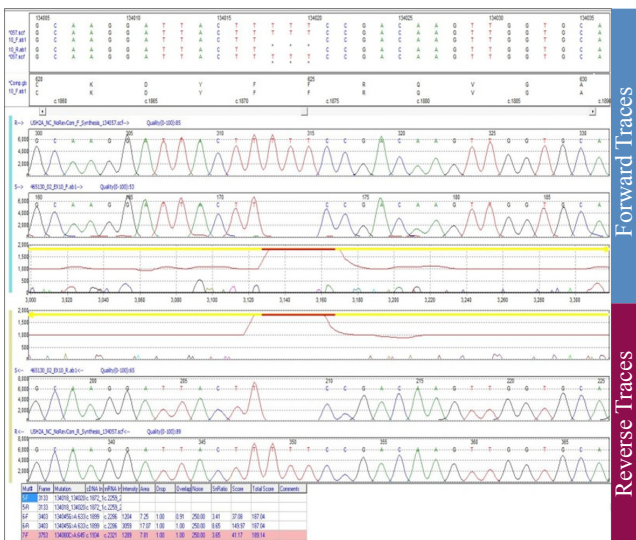


Figure 1: Example of a Homozygous Deletion. The mobility line turns red and a red bar is placed at the top of the mutation trace in the event of a homozygous indel event. A gap is placed in the sample trace indicating a 3 bp deletion (TTT).

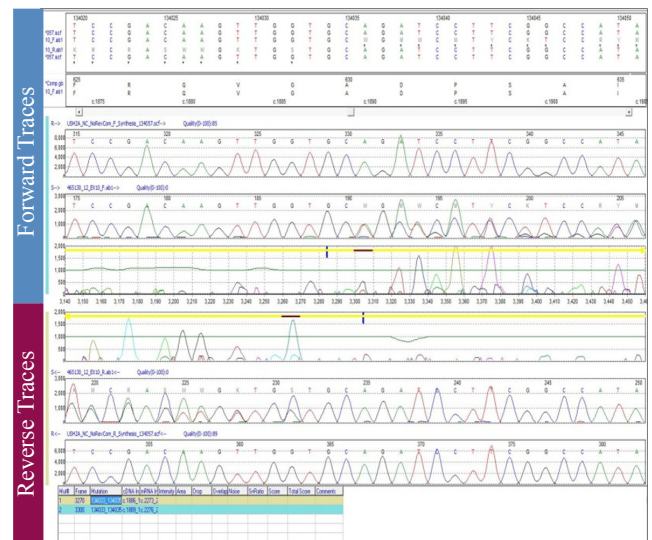


Figure 2: Example of a Heterozygous Duplication. The mobility line remains green and a brown bar is placed at the top of the mutation trace when a heterozygous indel is detected. The brown bar indicates a duplication of TGC.

Procedure

Data Entry

1. Add 1-Directional or 2-Directional sample files.
2. Add GenBank file or appropriate reference sequence files.
3. Click Process → Run to begin the analysis.


Analysis

Indels are reported in the Mutation Report, and are identified by horizontal bars at the top of the mutation trace in the Graphical Analysis Display (GAD). Homozygous indels are indicated by a red bar (figure 1) and heterozygous indels are located by a brown bar in the mutation trace (figure 2). The mutation trace is displayed in the third panel of the figures. The first panel is the reference trace synthesized from the GenBank file, and the second panel is the sample trace.

Results

Mutation Surveyor software uses a multi-step alignment algorithm to detect deletions, insertions, and duplications from Sanger sequence traces. Insertions and deletions are found by monitoring the mobility of the sample DNA fragments in comparison to the reference. The software's indel detection is not affected by basecalling error when DNA migration times of sample and reference traces are correlated.

Homozygous indels, as shown in figure 1, are identified when the middle line of the mutation trace turns from green to red in the GAD. An upwards peak marks the location of the deletion and a downwards peak marks the location of an insertion.

While heterozygous indels do not show a red line in the mutation trace of the GAD (see figure 2), a red line is observed in the Heterozygous Indel Detection tool. This tool automatically deconvolutes sample traces when a heterozygous indel event has occurred. For further analysis, open the Heterozygous Indel Detection tool (figure 3). Click on the  icon in the software's main toolbar or by double-clicking in the desired mutation cell in the various reports.

A is the reference trace.

B is the sample trace.

The sample trace is deconvoluted into two portions following the point of the insertion or deletion: the conserved sector and mutation sector.


C is the conserved sector.

D is the mutation sector.

Mutation Surveyor then shifts the mutation sector to align to the reference trace. The software accurately detects the heterozygous insertion with a red arrow and the mobility line changes from blue to red in color.

E is the shifted mutation sector aligning to the reference.

F is the mutation trace depicting the location of the indel event with a red arrow.

There are factors that could cause the location of heterozygous indel to be detected at the wrong location such as discrepancies between the forward and reverse reading directions caused by repeat nucleotide sequences and cloning artifacts that create a deletion near the insertion site. In cases such as these, use the  run button of the Heterozygous Indel Detection tool. This option runs a more rigorous detection algorithm for deconvoluting the mixed trace and aligning the mutant trace to the reference trace.

Heterozygous indels may be edited by clicking on the Mutation button and modifying the starting position. Mutation Surveyor software will then recalculate the indel based on user constraints.

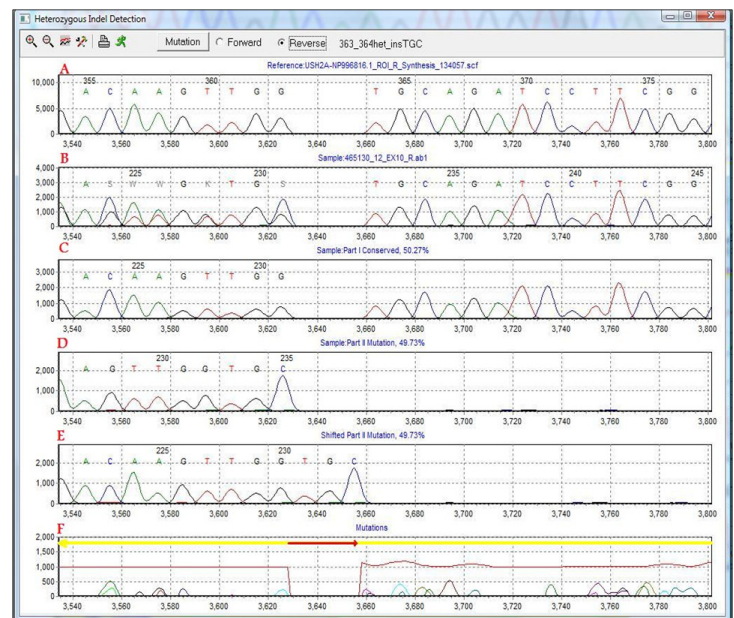


Figure 3: The Heterozygous Indel Detection Tool automatically deconvolutes the mixed mutation trace into two clean traces: the conserved trace (C) and the mutant trace (D). A Heterozygous insertion of TGC is indicated by the expansion of the reference trace (A), the arrow at the top of the mutation trace (F), and the red mobility line in the mutation trace.

A single sample trace can contain multiple heterozygous insertion/deletion events. In this case, click on the Mutation button, add a second mutation, and specify the location.

Notes

SoftGenetics' Mutation Surveyor software utilizes a unique physical trace to trace comparison (U.S. Patent No. 7,400,980) of the sequence traces and the reference traces rather than relying on the text basecall. By monitoring the migration time of the sample and the reference, Mutation Surveyor software can accurately detect heterozygous indels while minimizing false calls caused by text based comparison and alignment.

We recommend using a synthetic trace created from the GenBank text file as a reference since the synthetic traces will not contain any PCR artifacts. If reference sequence files are selected, it is important to input the correct reference trace in order for the software to detect Heterozygous Indels. It is recommended to use a parental sample as a reference. In the case of sample files including tandem repeats, the most appropriate reference trace would have the same number of repeats as the sample(s).

References

1. Hou et al. (2010) The D-allele of ACE insertion/deletion polymorphism is associated with regional white matter volume changes and cognitive impairment in remitted geriatric depression. *Neuroscience Letters*. 479: 262-266.
2. Wang et al. (2009) A six-nucleotide insertion-deletion polymorphism in the CASP8 promoter associated with risk and progression of bladder cancer. *Clinical Cancer Research*. 15: 2567
3. Zhang et al. (2009) A functional insertion/deletion polymorphism in the promoter region of the NFKB1 gene increases susceptibility for prostate cancer. *Cancer Genetics and Cytogenetics*. 191: 73-77.
4. Dong, C., Vincent, K., and Sharp, P. (2009) Simultaneous mutation detection of three homoeologous genes in wheat by High Resolution Melting analysis and Mutation Surveyor®. *BMC Plant Biology*. 9:143
5. Ellard et al. (2009) Semi-automated unidirectional sequence analysis for mutation detection in a clinical diagnostic setting. *Genetic Testing and Molecular Biomarkers*. 13:3: 381-386
6. Muram-Zborovski et al. (2010) NF1 exon 22 analysis of individuals with the clinical diagnosis of neurofibromatosis type 1. *American Journal of Medical Genetics*. 152: 1973-1978.
7. Estep et al. (2007) Mutation analysis of BRAF, MEK1 and MEK2 in 15 ovarian cancer cell lines: Implications for Therapy. *PLoS ONE*. 2(12); e1279
8. Valencia et al. (2013) Comprehensive mutation analysis for congenital muscular dystrophy: a clinical PCR-based enrichment and next-generation sequencing panel. *PLoS ONE*. 8(1): e53083