Variant Knowledge Base with External Annotation

Brent DiGiorgio, Kevin LeVan, ChangSheng, Jonathan Liu, Softgenetics LLC

Introduction

Mutation Surveyor software v5.0 includes a new Variant Knowledge Base designed to enhance the process of variant discovery and reporting in Sanger Sequence traces. The Knowledge Base can be configured to remember commonly made deletions such as false positives or artifacts as well as display external database annotation from common variant databases such as ClinVar, dbSNP, COSMIC, and dbNSFP. The new Graphic Analysis Display and mutation reporting options streamline the discovery process, allowing users to quickly identify variations of clinical significance and automatically eliminate false positives or sequencing artifacts from the mutation project.
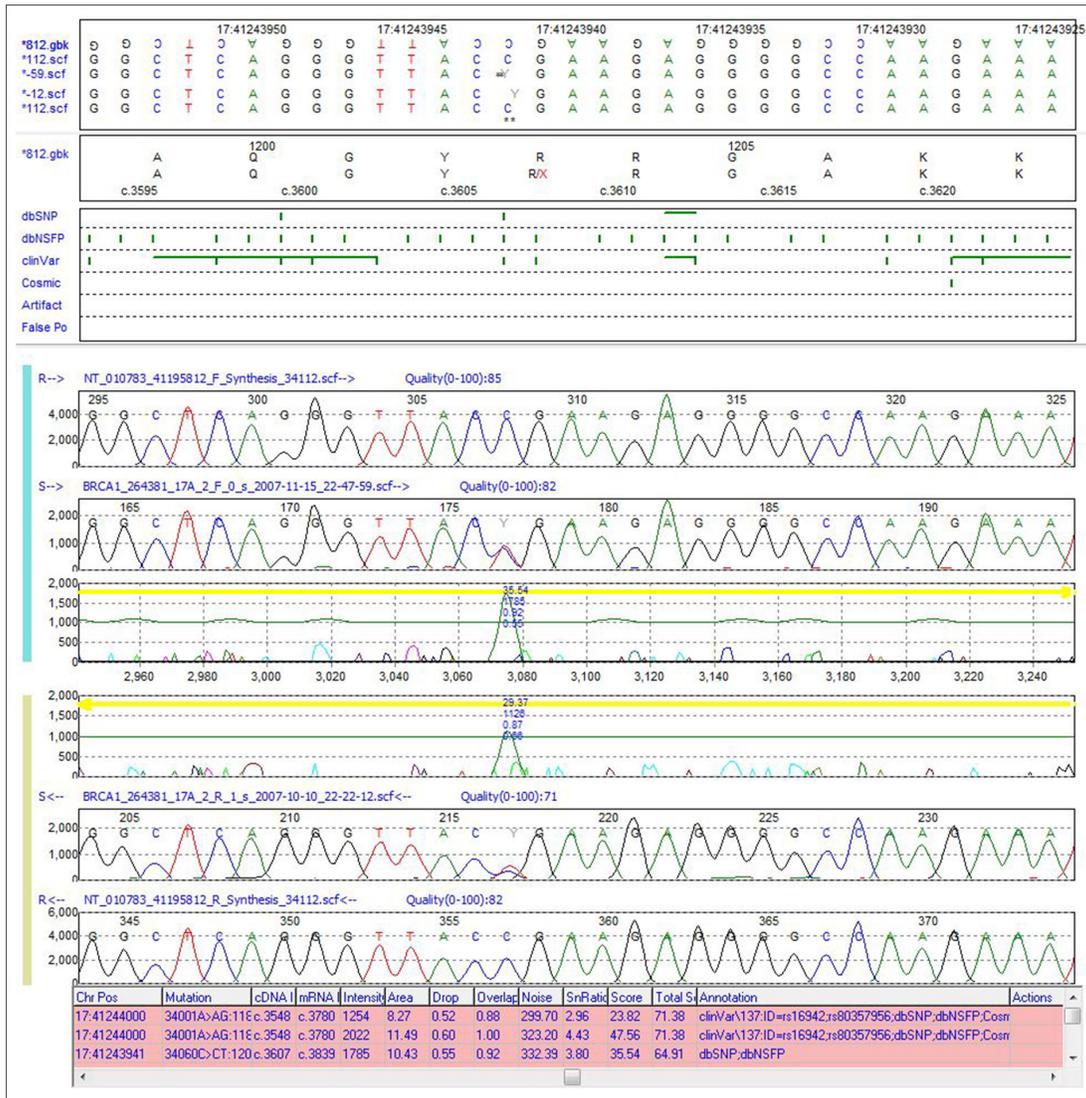

Figure 1: Graphic Analysis Display of Mutation Surveyor software showing chromosome position and database annotation in the mutation report at the bottom of the screen and the track pane at the top of the screen.

Procedure

Import Whole Human Genome Reference

1. Open Mutation Surveyor software by right clicking and selecting "Run as Administrator".
2. When Mutation Surveyor opens, go to Tools > Import Reference.
3. Follow the steps in the wizard to select and import a reference.

Import External Database Annotation

1. Navigate to Tools > Track Manager.
2. Select directory of reference file and select genome build.
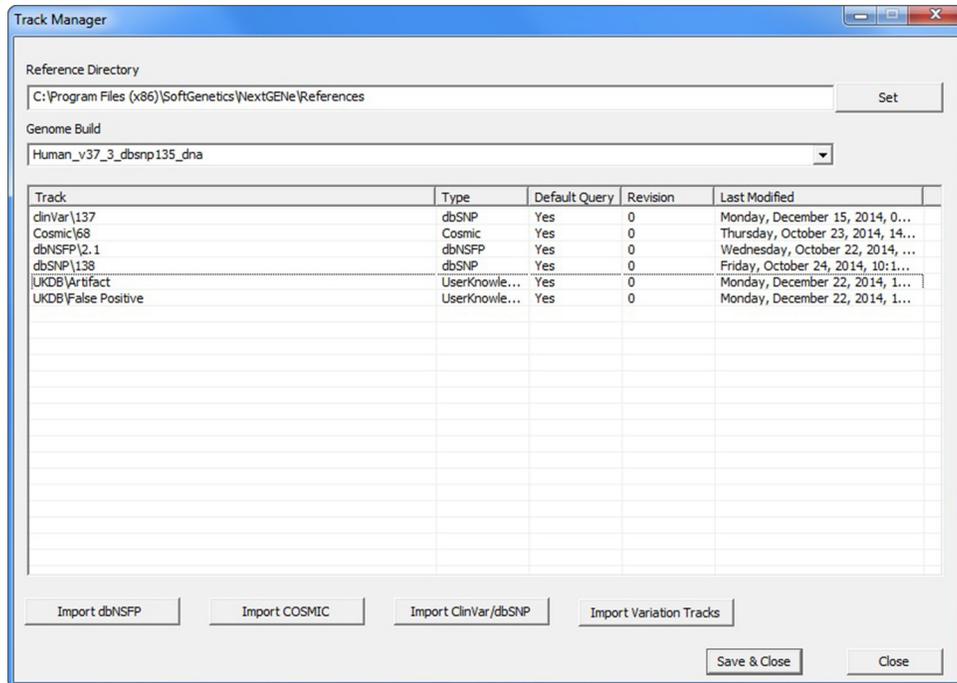3. Select external database icons to import database annotation.

Figure 2: The Track Manager of Mutation Surveyor allows users to select a reference genome and import external database annotation.

Add False Positive or Artifact Entries to Database

1. Right click mutation entry in mutation report and select "Delete Mutation".
2. Check the box to "Add to User Knowledge Database" and use the drop down option to label it as a False Positive or Artifact. Mutation Surveyor will automatically add this entry to the database.
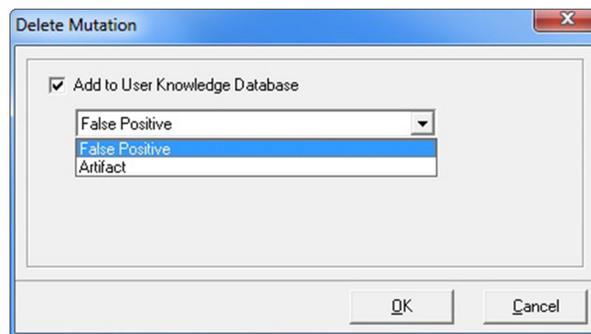

Figure 3: Mutation Surveyor software allows users to label deleted variants as False Positives or Artifacts and add them to a working database for future projects.

Query Databases in Mutation Project

   1. Navigate to Process > Query Reference Tracks to open a list of databases.

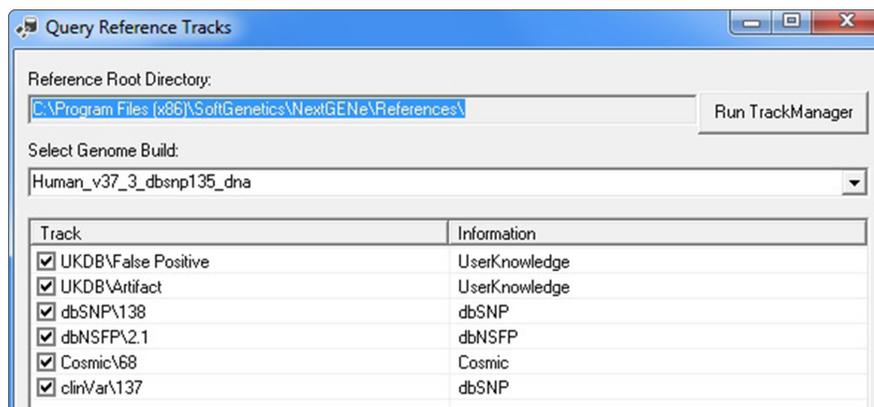   2. Check databases to query in current and future mutation projects.

Figure 4: Query Reference Tracks option showing the list of databases to be queried for mutation projects.

Results

The Graphic Analysis Display as shown in Figure 1 contains information related to sample data and database annotation including mutation information, chromosomal position, and database identifier.

Chromosome number and position are included in both the nucleotide sequence pane and mutation report for easy identification.

The Track Pane includes a list of imported databases along with their annotation. Green tick marks represent positions in the genome with variant annotation, including imported databases as well as custom entries of false positives and artifacts. To display statistics related to annotation, including the reported variant, dbSNP ID, COSMIC ID, and dbNSFP scores, hover the mouse pointer over the green tick mark.

The Mutation Report also includes an annotation column with relevant database information. The information displayed in this column can be customized by selecting the Tracks Setting icon ( ) at the top of the page.

The example in Figure 1 contains a variant located in the BRCA1 gene at chromosome position 41244000. The Track Pane displays tick marks for three external databases at this position: dbSNP, dbNSFP, and ClinVar. The annotation column is set to display all database information, including ClinVar version and rs ID.

Discussion

Mutation Surveyor software provides fast, accurate, and sensitive DNA variant detection from capillary sequencing. Along with variant detection, the Variant Knowledge Base feature of the software offers the ability to easily monitor chromosomal position, track false positives and artifacts, and identify positions of clinical significance. This database information is included in novel features in the Graphic Analysis Display, streamlining the process of variant discovery and reporting.