

Somatic Mutation Analysis from Whole Exome Sequencing Data with NextGENe® Software

Megan McCluskey, Jonathan Liu, John McGuigan, Ni Shouyong

Introduction

Somatic mutation analysis presents a challenge as these variants are often found at low frequencies, many times as low as 1-5%. Lowering variant calling thresholds to detect these variants can result in the increased reporting of false positives due to instrument sequencing errors. It's also necessary to differentiate the somatic mutations from non-somatic variants that are also present in the sample. The reduction of false positives is critical when the coverage for some captured exons is significantly lower than for others. False positives could also be caused by the sequence alignment due to the repetitive nature of the genome sequence.

A new Somatic Mutation Comparison Tool has been added to the NextGENe Viewer Comparisons menu in version 2.4.0. This specialized version of the Variant Comparison Tool allows the user to load three projects- a tumor project, a matched normal project (such as a blood sample), and a pooled project consisting of normal data from multiple sources. The tumor and normal projects are used to find somatic variants by filtering out variants that are shared in both projects, and the pooled project is used to eliminate artifacts due to library preparation and alignment by filtering out variants that occur in both the tumor and pool sample.

Procedure

The input for the Somatic Mutation Comparison Tool are project (*.pjt) files, created following the alignment and variant calling by NextGENe. The first step is to create the three individual projects – tumor, matched normal and pool. To create the pool project, multiple normal samples can be processed together to create a single pool project. The recommended number of normal samples for the pool is 4-5. The Mutation Filter settings should be adjusted for somatic mutation detection, with a low threshold for Mutation percentage, such as 5%, as shown in Figure 1 below.

Mutation Filter	<input type="checkbox"/> Use Original SNPs	<input checked="" type="checkbox"/> Except for Homozygous Indels	HomopolymerIndels
Mutation Percentage <=	5	5	5
SNP Allele Count <=	3	3	3
Total Coverage Count <=	10	10	10
Balance Ratios and Frequency <=	<input type="checkbox"/> 0.1 80 %	<input type="checkbox"/> 0.1 80 %	<input type="checkbox"/> 0.8 80 %

Figure 1: Suggested Mutation filter settings for somatic variant detection in whole exome sequencing data

After the analysis of all three projects is completed, the projects can then be loaded in the Somatic Mutation Comparison Tool.

1. Open NextGENe Viewer by going to File > Open NextGENe Viewer
2. Select Somatic Mutation Comparison Tool from the Comparisons menu in the Viewer.
3. Go to File > Load Projects to open the dialog where the three projects can be loaded.
4. Click the  icon to load the *.pjt file for each project type – tumor, matched normal and pool.
5. Adjust settings as needed.
 - a. Maximum Contamination – somatic variants may have up to this allele frequency in the normal project. This is to account for contamination of the normal sample with tumor sequence.
 - b. Number of Pooled Samples – The number of samples used in the pooled project. The Maximum Contamination setting is divided by the square root of this number and used to account for potential contamination in the pooled project.
 - c. Somatic Allele Count – The tumor sample must have at least this number of reads containing the variant.
 - d. Relative Directional Balance (T/N) – The forward/reverse balance ratio in the tumor project divided by the same ratio in the normal project must not exceed this value. This option can be deselected to skip filtering based on the relative balance. Selecting this option can filter out variants that are less directionally balanced in the tumor sample than in the normal sample.
 - e. Somatic Allele Frequency Ratio (T/N) – The ratio of the mutant allele percentage between the tumor and normal projects must be greater than this value.
 - f. Pooled Allele Count Ratio (T/P) – The ratio of the number of reads with the mutant allele between the tumor and the pool projects must be greater than this value. This option can be deselected to skip filtering based on the pooled allele count ratio.

Procedure (cont.)

- Click Mutation Report Filter/Display Settings and/or Tracks Filter/Display Settings to modify the settings for the report.
- Click OK to generate the comparison.

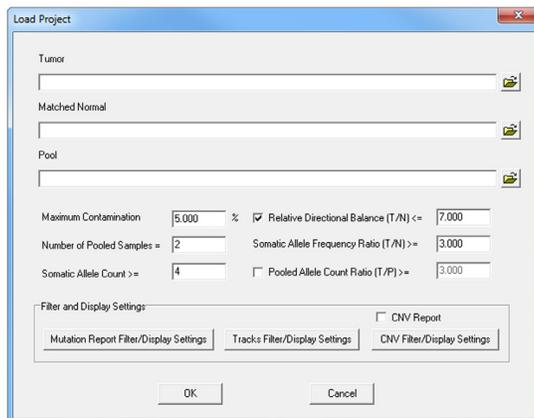


Figure 2: Load Project dialog of the Somatic Mutation Comparison Tool including settings

Results

The results of the comparison are shown in a table listing all variants that passed the filters **significantly reducing the number of false positives - tests indicate up to a 70% reduction**. The Somatic Mutation Comparison Tool results view is interactive. You can double click on any variant in the table to view the aligned reads at that position in all projects. You can also search by gene name or chromosome position. The report can also include information from variant databases, such as COSMIC and dbNSFP, which contains population frequencies from 1000 Genomes and Exome Variant Server. Databases can be imported through NextGENe's Track Manager Tool, and can then be displayed in the report and/or used for filtering the report.

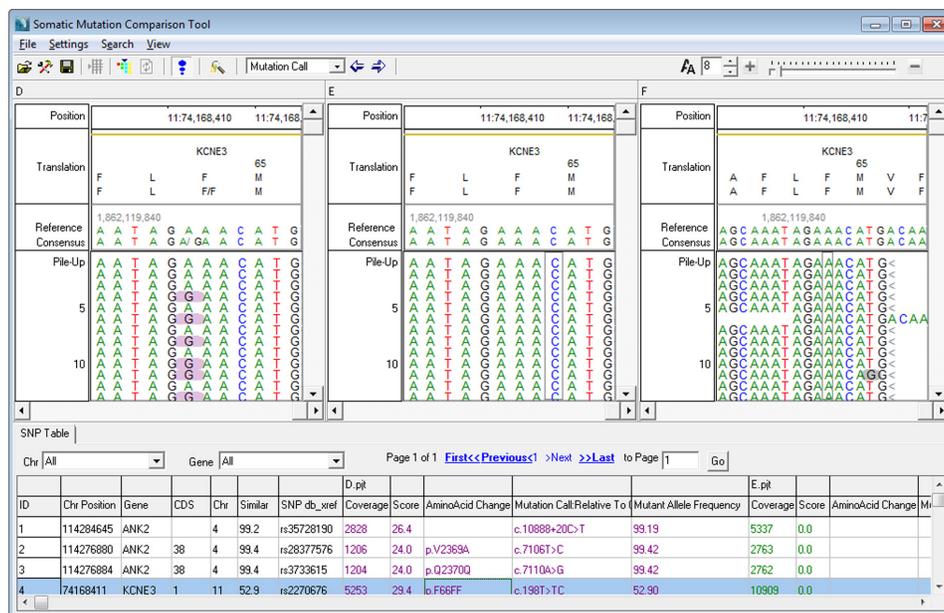


Figure 3: Somatic Mutation Comparison Tool results

In tests using 5 pooled controls the number of false positive calls has shown a reduction of up to 70%. Many of the remaining false positives were caused by common variants and systematic errors that appear in the tumor sample but were not called in the control sample due to low coverage. Increasing the coverage of the pooled normal samples will reduce the artifacts and maintain the specificity to the somatic mutation allele.