

Highly Sensitive Somatic Mutation Detection from Biopsy Samples with Deep Sequencing using NextGENe® Software

Kevin LeVan, John McGuigan, Shouyong Ni, CS Jonathan Liu

Introduction

A biopsy is often ordered by a physician in order to more closely examine a tissue sample. The tissue is usually examined by a pathologist under microscope to determine if the tissue is cancerous, and the sample may be used for molecular variant detection to predict responses to possible therapies. Modern sequencing technologies enable liquid biopsies, allowing scientists to detect circulating DNA fragments (130 bp long on average) in the blood stream which are released by the breakdown of cancerous cells.¹ PCR is performed to selectively amplify DNA from regions of interest and sequenced on next-generation sequencing instruments and the same can be done for a control sample (cheek swab, white blood cells, etc) from the same patient.

The key difficulty in this analysis is the low frequency (1-10%) of variants which need to be detected. Despite the challenges, this liquid biopsy approach has been used to monitor the effects of cancer treatment over time using patient specific mutations in BRAF, NRAS, cKIT, TERT, and ALK genes during treatment.² The approach, recommended in some studies, was used by NextGENe software to perform variant comparison between a sample and control pair under the same experimental.³ Multiple criteria are useful for reducing false positives, including requiring a minimum variant read count of 15, balanced forward and reverse coverage ratio (<3), and absence of the variants in a control sample derived from white blood cells.

Procedure

1. Align sequence data and call variants using NextGENe for both sample and control data. Settings are shown in Figure 1.
2. Open the Somatic Mutation Comparison Tool in the NextGENe Viewer (found in the “Comparisons” menu) to compare the mutations between the test sample and control sample.

The screenshot shows the 'Project Wizard - Alignment' dialog box. On the left is a vertical sidebar with buttons for 'Application', 'Load Data', 'Condensation', 'Assembly', 'Alignment', and 'Post Processing'. The main area contains the following settings:

- Alignment:**
 - Reads: Allowable mismatched bases: 0 (0-2)
 - Allowable ambiguous alignments: 20
 - Seeds: 37 bases, move step: 1 bases (with 'Inspect input files' button)
 - Allowable alignments: 500 (1-1000)
 - Overall: Matching base percentage >= 90 (with 'Detect large indels' checkbox)
- Sample trim:**
 - Select sequence range: from 1 bases to 30 bases (checkboxes for 'from' and 'to')
 - Hide unmatched ends (checkbox)
- Mutation filter:**
 - Use original SNPs (checkbox)
 - Except for homozygous Indels HomopolymerIndels (checkboxes)
 - Mutation percentage <= 1
 - SNP allele count <= 15
 - Total coverage count <= 100
 - Balance ratios and frequency <= 0.2 (checkbox) / 80 %
- File type:**
 - Load assembled result files (checkbox)
 - Load SAGE expression data (checkbox)
 - Extract bases from: 2 bases to: 17 bases
 - New sequence coverage minimum: 20
 - Load paired reads (checkbox)
 - Library size range: from 0 bases to 600 bases
 - 454 Sequence: (text field)
 - Save matched reads (checkbox)
 - Highlight anchor sequence (checkbox)
 - Ambiguous gain/loss (checkbox)
 - Detect structural variations (checkbox)
 - Mismatch: 0.1 length and 20 bases (checkboxes)
 - Default Settings (button)

At the bottom are buttons for 'Save Settings', 'Load Settings', '<< Back', 'Next >>', 'Cancel', and 'Finish'.

Figure 1: Sequence alignment settings for high coverage targeted sequencing. The minimum coverage is set to 100x, minimum SNP allele count is set 20x, minimum mutation percentage is 1%, and minimum balance ratio is set to 0.2 for tumor mutation detection. The minimum mutation percentage for the normal control sample is set to 5% to tolerate contamination from where tumor cells may have metastasized.

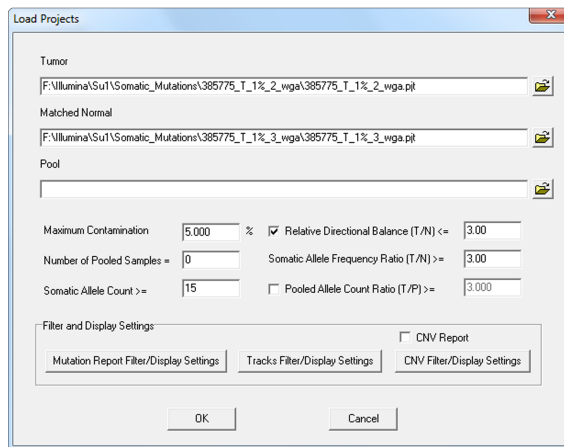


Figure 2: Settings for somatic mutation detection of PCR products.

The input for the Somatic Mutation Comparison Tool are project (*.pjt) files, created following alignment and variant calling by NextGENe. The pooled sample is not needed for very high coverage targeted sequencing (>1000x), but is more useful for typical exome sequencing projects which usually have a lower level of coverage (around 150x). The Comparison Options should be adjusted for somatic mutation detection, with a low threshold for Mutation percentage, such as 1%, as shown in Figure 1. The Maximum Contamination of the normal project (set to 5% in Figure 2) will allow detection of mutations while tolerating tumor contamination of the normal sample with variants up to this frequency. The “Check Allele Counts for Negative Mutations” option in the Mutation Report Settings “Statistics” tab should be checked in order for the software to examine counts for each allele.

- **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. It is set to 5% of contamination in normal project.
- **Number of Pooled Samples** – The number of samples used in the pooled project. This is useful when the coverages of some regions are lower. It is not needed for targeted PCR products.
- **Somatic Allele Count** – The tumor sample must have at least this number of reads containing the variant.
- **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. The sequence reads in PCR-based targeted sequencing are often well balanced (1 to 3). 1 is perfectly balanced and 3 is about 3:1 forward/reverse or reverse/forward.
- **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. This means the normal must have lower background.
- **Pooled Allele Count Ratio (T/P)** – This setting is used when a pooled background sample is loaded.

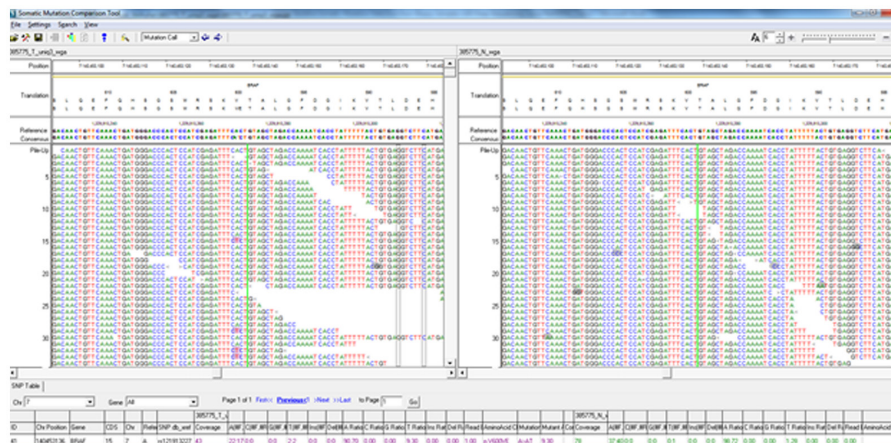


Figure 3: The BRAF, V600E, mutation was found on chromosome 7 in the patient biopsy sample on the left at 4.9% tumor mutation. The white blood cell control sample on the right did not show this mutation even though there is an error.

Discussion

False negatives and false positives. The release of DNA from tumor cells into the blood stream may be limited depending on the type of tumor, the position of the tumor, and the size of the tumor. Liquid biopsies that start from a higher volume of blood may be more successful in detecting low concentrations of free DNA in the blood. False positives can be minimized by using high coverage sequencing with a detection threshold of 1%, and a narrow range of allowed forward:reverse balance ratios. In addition, PCR artifacts can attribute to some false positives when detection sensitivity is set as high as 0.1-1%. A small portion of reads (less than 1-2%) shorter than the amplicon may have errors, especially within the first and last 20bp of each read, and accuracy can be improved when these short reads are removed from the analysis.

Drug response detection for longitudinal studies. Experiments should be performed under similar experimental conditions to allow comparison of data from different periods with less noise. The data have shown that the amount of the circulating DNA is highly correlated with the progress of tumor measured with radiography as well as progress duration. The quantitative measurement of progress from 3% to 5% frequency of a mutant allele (NRAS Q61R) over a period of three months has shown significant correlation with the days after 1st drug treatment ($R^2 > 0.99$).²

References

- 1 Bettegowda, Chetan, et al. "Detection of circulating tumor DNA in early-and late-stage human malignancies." Science translational medicine 6.224 (2014)
- 2 Lipson, Evan J., et al. "Circulating tumor DNA analysis as a real-time method for monitoring tumor burden in melanoma patients undergoing treatment with immune checkpoint blockade." Journal for immunotherapy of cancer 2.1 (2014)
- 3 Jones, Siân, et al. "Personalized genomic analyses for cancer mutation discovery and interpretation." Science translational medicine 7.283 (2015)

**Research Use Only*