3582

Copy Number Estimation from Targeted and Shallow Sequencing in Cancer Samples

Zhiwei Che, Westley Sherman, Andrea O'Hara, Raja Keshavan, Viren Wasnikar, Soheil Shams

BioDiscovery, Inc., 715 N. Douglas St., El Segundo, CA 90245

Abstract

Next-generation sequencing (NGS) is mainly used to obtain sequence variants (SNVs). However, getting copy number results from NGS has gained momentum in both research and clinical applications. Here we report an algorithm, BAM (MultiScale Reference), for deriving copy number results from targeted panel NGS and shallow WGS data, as well as from WGS and WES with normal depth of coverage. This method builds a reference file out of a pool of BAM files that are either from normals or from unrelated experimental samples, then makes copy number and loss of heterozygosity (LOH) calls for each sample, based on the logRatios and B-allele frequencies (BAFs). As part of the Nexus Copy Number software, it gives scientists a straightforward way to analyze NGS data for copy number events in a graphical interface.

Materials and Methods

Here we introduce the BAM (MultiScale Reference) algorithm, currently in Nexus Copy Number 9.0, to function with shallow and targeted sequencing data, as well as WGS and WES, by introducing a novel dynamic binning approach. This approach uses a Hidden Markov Model to segment the genome into "target" areas using the reads in targeted regions and the "backbone" areas using the off-target reads and additional areas. It uses coarse binnings in the "backbone" areas that provides copy number base line as well as large copy number events and uses fine binnings in "target" areas to provide high resolution copy number detection in targeted regions. Shallow WGS data and targeted panel NGS data, as well as WES with normal depth of coverage, were used for the testing. The results were compared with those from microarray and/or other algorithms in Nexus Copy Number.

Conclusions

The BAM (MultiScale Reference) method described here is of special value, because it works with a pool of random experimental samples to build the reference, not necessarily requiring the "real" normal samples. The "target" and the "backbone" areas present a contrast for the copy number profiling. Most importantly, the "backbone" areas in cancer samples can be used as diploid regions for ploidy adjustment, which is critical for correct cancer copy number analysis. The good news for scientists is that it is incorporated nicely in the Nexus Copy Number software, which provides a graphical software solution to get copy number and LOH calls from WGS, WES, shallow sequencing, and targeted panel NGS data, in addition to its support for virtually all microarray platforms.

Results

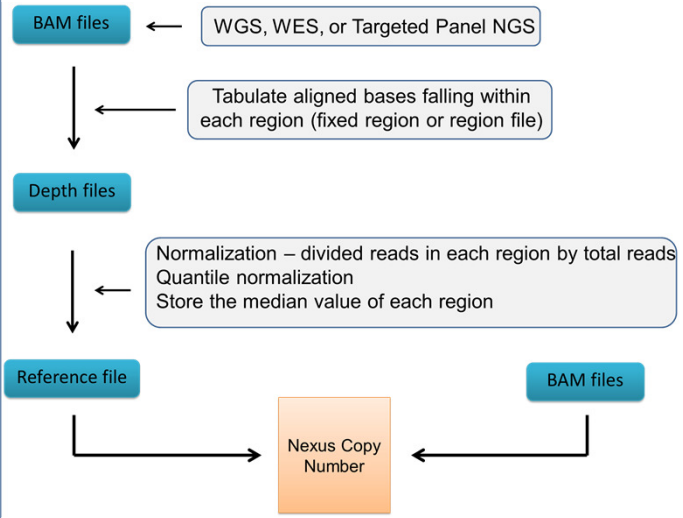

Figure 1. Summary of the method used to estimate copy number from NGS data. A reference file is created from pooled reference BAM files. Individual BAM files of the experimental samples are then loaded into Nexus Copy Number against the reference file.

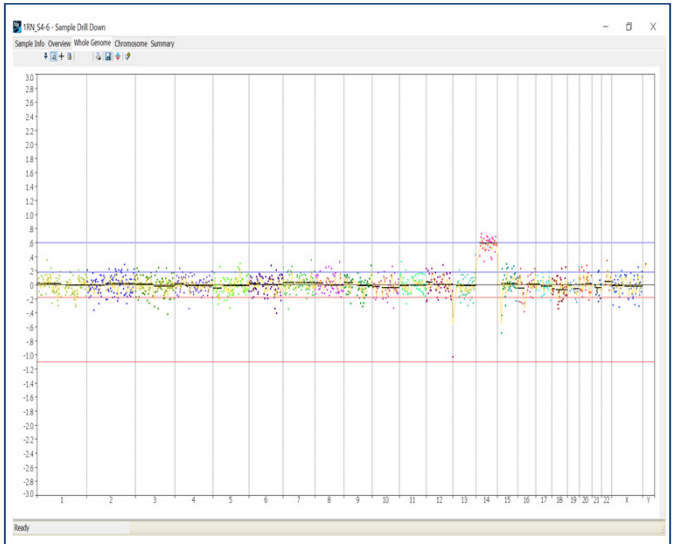

Figure 2. A shallow whole genome sequencing (WGS) sample analyzed by the BAM (MultiScale Reference) method shows a duplication on chromosome 14 in this whole genome logRatio plot. Each point in the plot is a pseudo-probe that represents a certain genomic bin (region).

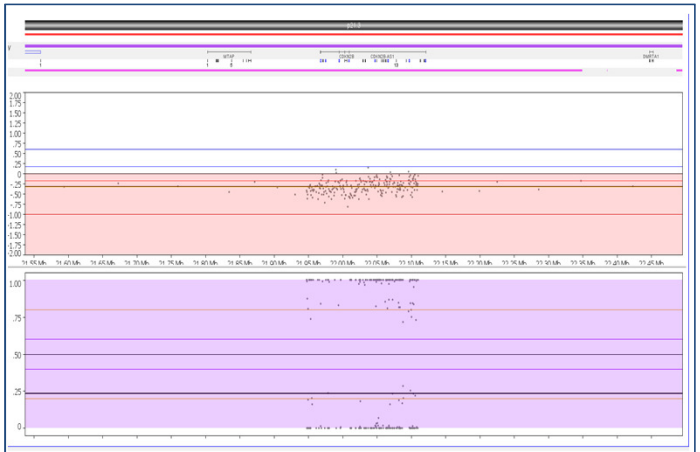

Figure 3. A targeted panel NGS lung cancer sample, showing a copy loss in the CDKN2A and CDKN2B genes. Many pseudo-probes are present inside the two genes due to the very high coverage in the target region. The off-target regions outside these two genes have fewer pseudo-probes due to lower coverage, also showing CN loss.

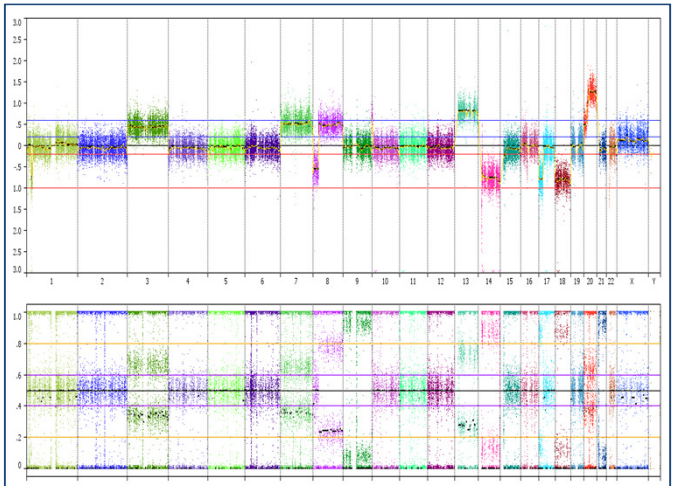

Figure 4. The method also deals with normal WGS/WES (e.g. 30x) data. Copy number profile of a TCGA COAD whole exome sequencing (WES) sample shows both logRatios and B-allele frequencies (BAFs).