Detection of genome-wide copy number variation using the Applied Biosystems™ Axiom™ Genotyping Solution

Ali Pirani, Ram Varma, and Mohini A Patil
Thermo Fisher Scientific, Central Expressway, Santa Clara, California, USA, 95051

ABSTRACT

Copy number variations (CNVs) have been implicated in both disease phenotypes and phenotypic variation associated with quantitative traits in plant and animal species. Software tools available from Thermo Fisher Scientific offer a valuable resource for evaluation of the impact of CNVs with respect to variation in plant and animal health and production traits. Applied Biosystems™ Axiom™ Analysis Suite Software now has enhanced CNV capabilities that utilize intensity and genotypes to calculate log2 ratios and B-allele frequencies (BAFs) from genotyping data for detailed analysis and visualization.

This platform has the dual ability to detect CNVs in targeted genomic regions and to genotype single nucleotide polymorphisms (SNPs) across the whole genome using a single assay. CNV analysis methods include (a) whole genome de novo analysis for discovery and (b) fixed region analysis when breakpoints of CNV regions of interest are known *a priori* and there is little breakpoint variability from sample to sample. In CNV discovery analysis, CN states are determined by a Hidden Markov Model (HMM) implementation. Breakpoints are discovered and CN segments are labeled. Fixed region analysis uses an optimized multi-sample clustering algorithm to assign CN states to each region in each sample.

These software tools allow the user to perform complex copy number analysis utilizing both methods. This results in superior analytical sensitivity and specificity for known small regions, while enabling discovery for larger regions and across the whole genome. Examples of copy number gains and losses are presented here.

INTRODUCTION

Applied Biosystems™ Axiom™ custom and catalog genotyping arrays offer innovative capabilities in genotyping applications for quantitative trait loci discovery, marker assisted selection, and genomic selection. Copy number variations (CNVs), or alterations in the number of copies of a stretch of a genome, are important in several areas of plant and animal research. CNVs can manifest themselves as copy number (CN) gains or losses, and have been associated with animal physical attributes and plant phenotypic diversity. CNVs have been implicated in diseases and traits in nonhuman species such as canine, bovine, porcine, poultry, and several diploid and polyploid plants. Therefore, in addition to genotyping single nucleotide polymorphisms (SNPs) and insertions or deletions (indels), Axiom arrays are designed to detect CN changes and allelic imbalances such as loss of heterozygosity (LOH). Axiom arrays can be used for targeted CNV applications or whole-genome CN discovery applications in animals and plants.

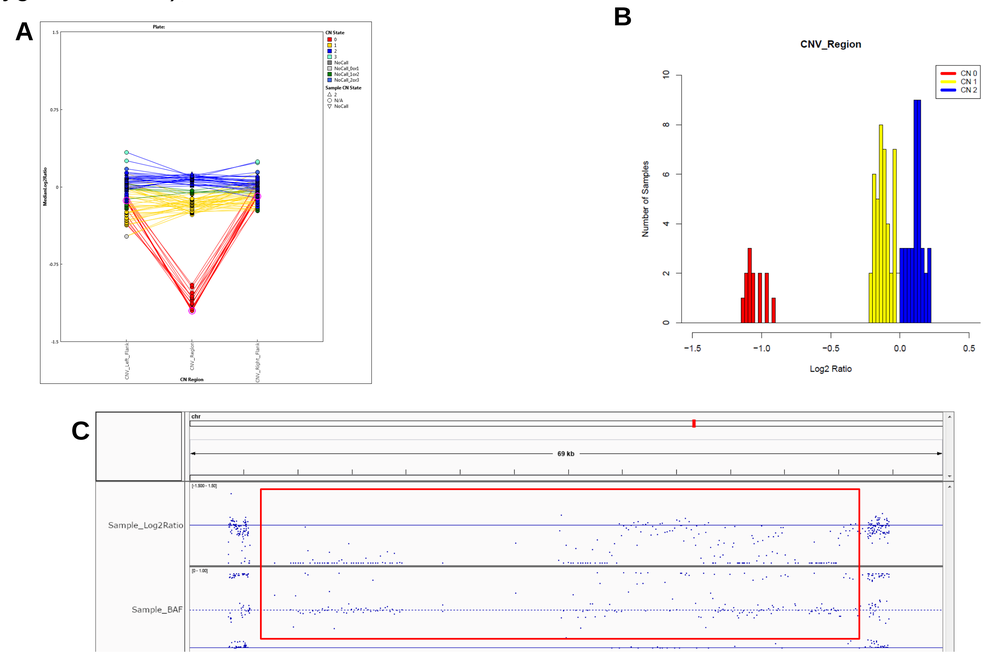
CNV ANALYSIS

The $\log_2$ ratio and BAF, which form the basis for CNV calling, are calculated directly within the Axiom Analysis Suite application and used with the workflows established for CN analysis. The $\log_2$ ratio is the $\log_2$ of the ratio of signal intensity of a probeset to reference total intensity for the same probeset. The reference total intensity is an estimate of the total A and B allele intensities for the probeset, representing the normal diploid state at that location. The BAF is a measure of the heterozygosity at any location and is calculated for each probeset: Raw BAF = (signal intensity of B allele)/(signal intensity of A allele + signal intensity of B allele). Probesets used for CN calculations include those at polymorphic and nonpolymorphic markers. Probeset selection for CN analysis may be based on sequence homology and signal response to CN changes.

Fixed-region analysis

Fixed-region analysis is one of two methods for performing CN analysis in Axiom Analysis Suite or APT (Applied Biosystems™ Array Power Tools). This method is used when the breakpoints of CN regions of interest are known from publications or prior work. Fixed-region CN analysis requires designing probesets for a specific gene or a specific region within the genome (e.g., a CN analysis of specific exons of the canine FGF3 gene is achieved by designing multiple probesets across the FGF3 gene). By saturating the targeted region with multiple probesets, one is able to get higher resolution of the CN changes within the region. Probesets used for CN analysis can interrogate nonpolymorphic markers or SNPs in the region. Example results are shown in Figure 1.

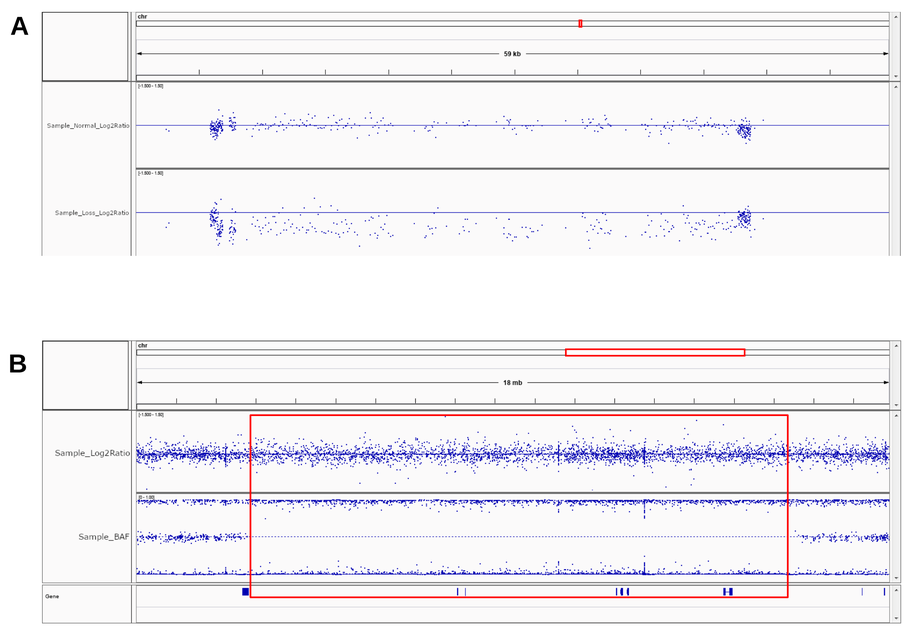
Figure 1. Example plots of a fixed-region CNV analysis. **a** Plate-level visualization from Axiom Analysis Suite. Each point is a sample. The CNV region is the middle column. The left and right columns are the left and right flanks, respectively. **b** Region-level histogram view for the same CNV region. **c** Sample-level visualization for the same region from Integrative Genomics Viewer (IGV). Output formatted for IGV can be generated using APT. The region of interest is indicated with a red box, with flanks both upstream and downstream. The flanks are CN2, while the region is mostly CN0 (homozygous deletion), with some additional variation.



Whole-genome de novo analysis

Discovery or *de novo* analysis is another method for performing CN analysis in Axiom Analysis Suite or APT. This method is used when CN breakpoints are not known and must be determined. CN states are determined by implementation of a hidden Markov model (HMM). Breakpoints are discovered and CN segments are labeled by states. Regions of interest are specified as inputs based on intent at the time of array design. State transition probabilities, model parameters, and state priors may also be specified as inputs. Example visualization of results from a discovery analysis in IGV is shown in Figure 2.

Figure 2. Example plots of a discovery CNV analysis. Sample-level visualization for a discovered region from Integrative Genomics Viewer (IGV). **a** The first track is the $\log_2$ ratio for a sample with normal copy number (CN2) in this region. The second track is the $\log_2$ ratio for a sample with a copy number loss (CN1) in this region. **b** Example of a discovered 14 Mb LOH region. Note the presence of genes in this region.



CONCLUSION

The Axiom platform enables copy number analysis with two methods: the Fixed Region method for known small regions; and the Discovery *de novo* method enables detection of events for larger regions and across the whole genome.

ThermoFisher SCIENTIFIC