Microarray analysis

# Copy number variation analysis using Axiom arrays

## Introduction

Copy number variations (CNVs) are structural changes of DNA including gains or losses that account for significant variation among human genomes [1]. They have been associated with increased risk for various human diseases, abnormalities, conditions, and developmental disorders [2]. CNVs also contribute to trait variation in nonhuman species such as canine, bovine, porcine, poultry, and several diploid and polyploid plants [3,4]. Therefore, in addition to genotyping single nucleotide polymorphisms (SNPs) and insertions or deletions (indels), Applied Biosystems™ Axiom™ arrays are designed to detect CNVs and allelic imbalances such as loss of heterozygosity (LOH). Applied Biosystems™ Axiom™ Analysis Suite 5.2 software provides the tools and visualizations needed to perform CNV analysis in regions with predefined boundaries (fixed-region analysis) and across the entire genome (discovery analysis).

## Copy number analysis

CNVs of test samples are detected using computed $\log_2$ ratios and B allele frequencies (BAFs) for probesets interrogating individual markers across the genome. The $\log_2$ ratio is the log-transformed ratio of the test sample signal intensity relative to a reference total intensity for the same probeset. BAF is a measure of heterozygosity, calculated for each probeset as the ratio of B allele signal intensity relative to the total signal intensity. The reference total intensity is an estimate of the total A and B allele intensities for the probeset. This reference value is calculated as the median total intensity for that probeset across reference samples.

Generating a reference in genomic regions where CNVs are common poses a special challenge. For example, in the human *GSTM1* gene, most individuals in a population may not be diploid. In such a region, the reference for probesets should be generated from carefully selected diploid samples. Axiom Analysis Suite 5.2 offers new workflows that (1) enable initial reference creation using a well-characterized set of samples and (2) identify new diploid samples to create a more robust reference for human pharmacogenomics and blood typing applications [5].

The analysis software offers two CNV analysis methods: (1) fixed-region analysis when breakpoints of CNV regions of interest are known a priori and there is little breakpoint variability from sample to sample and (2) discovery analysis to detect copy number (CN) changes across the whole genome. Fixed-region analysis provides superior analytical sensitivity and specificity for known small regions [6].

QC metrics used to select samples for analysis include MAPD, the median of absolute values of pairwise differences between $\log_2$ ratios of adjacent markers, and wavinessSD, a global measure of $\log_2$ ratio variation that is insensitive to short-range differences.

## Fixed-region analysis

Fixed-region CN analysis uses a novel algorithm, CNVMix, that calculates median $\log_2$ ratios for each region, performs multi-sample clustering analysis for each region, and assigns CN states to each sample. Probesets are selected in fixed regions at the time of array design, and the mapping information is stored in library files. Example results are shown in Figure 1.

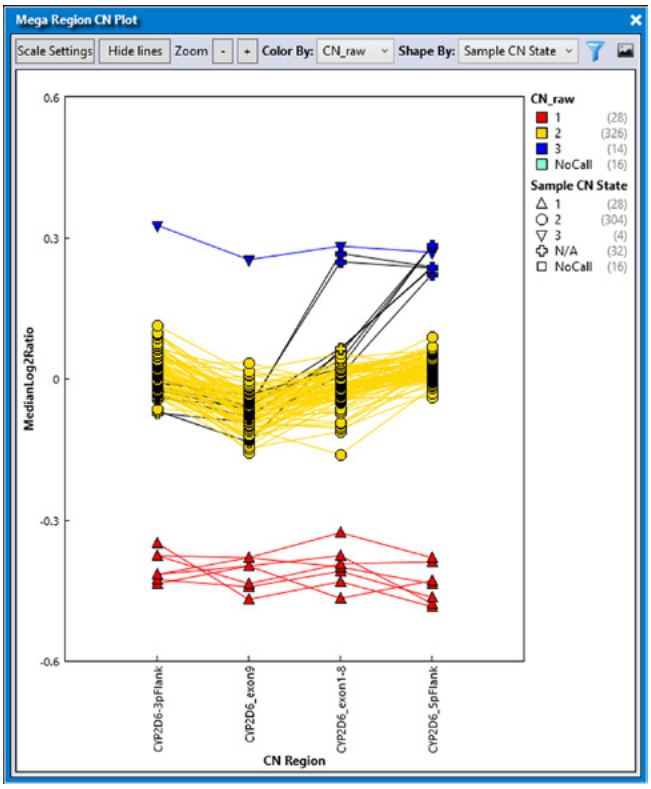


**Figure 1. Copy number calls for four *CYP2D6* regions for 92 samples from the International HapMap Project.** Samples were run on the Applied Biosystems™ Axiom™ Precision Medicine Research Array (Cat. No. 902981). Changes in copy number state are shown along the gene in many samples. Mega Region Plots were generated using Axiom Analysis Suite.

## Discovery analysis

In discovery CN analysis, CN states are determined by a hidden Markov model (HMM) applied to $\log_2$ ratios of ordered markers on genomic regions of interest such as entire chromosome arms. Probesets for discovery analysis are selected at the time of array design. Breakpoints are discovered and the CN states of segments are labeled by the HMM algorithm. Results from a discovery CN analysis are shown in Figure 2.

LOH regions are those where SNPs do not show any heterozygosity (Figure 3). Probesets at markers with high minor allele frequencies are used to detect regions of LOH. In agrigenomics applications, such regions may also indicate absence of heterozygosity due to inbreeding rather than LOH.

For more details on genotyping using Axiom arrays, please refer to the Axiom Genotyping Solution Data Analysis Guide (Pub. No. MAN0018363). For more details on CNV analysis using Axiom arrays, please refer to our Axiom copy number analysis technical note (Pub. No. COL32811).

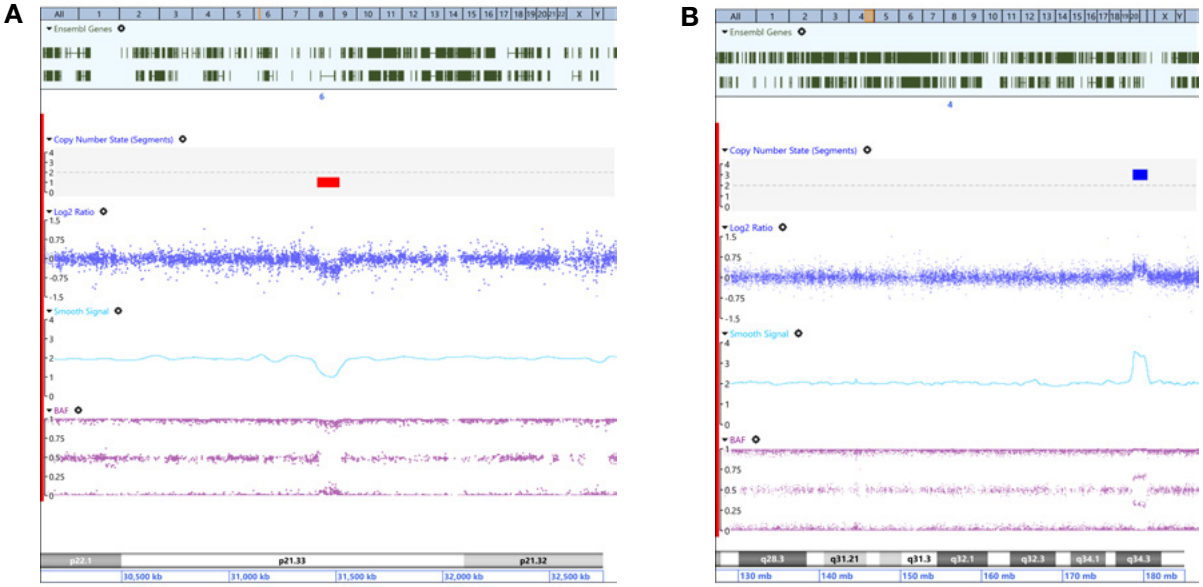


**Figure 2. Visualization of copy number changes. (A)** Single-copy loss on chr6:31,386,000–31,484,000 in a sample assayed using the Applied Biosystems™ Axiom™ Precision Medicine Diversity Research Array (Cat. No. 951962) visualized using Whole Genome View in Axiom Analysis Suite 5.2. Points on the $\log_2$ ratio and BAF tracks represent values measured by probesets. The $\log_2$ ratio track shows a depression from normal state in the deletion. The BAF values show only two bands representing alleles A and B in the deletion, while the flanking normal diploid regions show three bands representing AA, AB, and BB alleles. **(B)** Single-copy gain on chr4:177,197,000–178,966,000 in a sample assayed using the Axiom Precision Medicine Diversity Research Array (Cat. No. 951962) visualized using Whole Genome View in Axiom Analysis Suite 5.2. Points on the $\log_2$ ratio and BAF tracks represent values measured by probesets. The $\log_2$ ratio track shows an elevation from normal state in the duplication. The BAF values show four bands representing AAA, AAB, ABB, and BBB alleles.

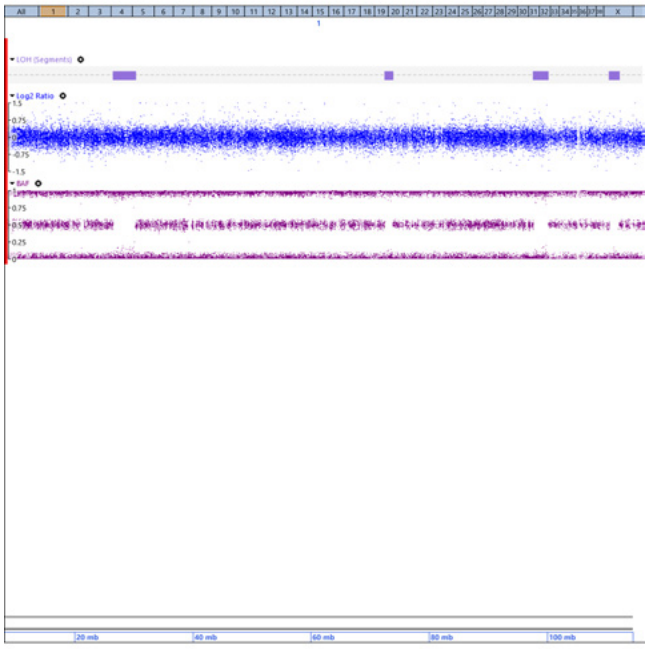

**Figure 3. Example data showing LOH.** Four LOH regions on chromosome 1 are observed in a canine sample assayed on the Applied Biosystems™ Axiom™ Canine HD Array (Cat. No. 550869). There are no copy number changes. The BAF track shows absence of heterozygosity (no AB alleles) in the four regions.

## Verification

CNV algorithms and workflows were experimentally verified using Axiom arrays for fixed-region and discovery CN analysis.

## Fixed-region analysis

Algorithm verification used data from the Axiom Precision Medicine Diversity Research Array (Cat. No. 951962). Fixed-region analysis across 10 regions of interest was performed on 6 plates of samples including an Axiom training plate. There were 445 known CN0 events in this verification data set, all of which were detected. No additional CN0 events were detected. There were 907 known CN1 events in this verification data set, of which 906 events were detected. No additional CN1 events were detected. There were 201 known CN3 events in this verification data set, of which 199 events were detected. Two additional CN3 events were detected. Across all copy number states, analytical sensitivity ranged between 82% and 100% and the analytical positive predictive value ranged between 86% and 100%.

## Discovery analysis

Algorithm verification used data from the Applied Biosystems™ UK Biobank Axiom™ Array (Cat. No. 902502) [7]. Whole-genome CN discovery analysis was performed on three plates of samples from the International HapMap Project. There was only one known CN0 event in the data set, and this event was detected. There were 79 known CN1 events in the verification data set, of which 72 events were detected. One additional previously unknown CN1 event was also detected. There were 173 known CN3 events in the verification data set, of which 154 events were detected. Twenty-eight additional previously unknown CN3 events were also detected. Overall, CN1 events were detected with analytical sensitivity of >90% and an analytical positive predictive value of >80%. CN3 events were detected with analytical sensitivity of >80% and an analytical positive predictive value of >70%.

### References


1. Conrad D, Pinto D, Redon R, et al. (2010) Origins and functional impact of copy number variation in the human genome. *Nature* 464:704–712.
2. Kendall KM, Rees E, Escott-Price V et al. (2017) Cognitive performance among carriers of pathogenic copy number variants: analysis of 152,000 UK Biobank subjects. *Biol Psychiatry* 82:103–110.
3. Strillacci MG, Cozzi MC, Gorla E et al. (2017) Genomic and genetic variability of six chicken populations using single nucleotide polymorphism and copy number variants as markers. *Animal* 11:737–745.
4. Berry D, O'Brien A, O'Donovan J et al. (2018) Aneuploidy in dizygotic twin sheep detected using genome-wide single nucleotide polymorphism data from two commonly used commercial vendors. *Animal* 12:2462–2469.
5. Axiom Analysis Suite 5.2 User Guide.
6. Varma R, Aull K, Bruckner C et al. (2019) Detection of copy number variations in blood typing research and ADME genes using Axiom assays. ASHG Annual Meeting. http://assets.thermofisher.com/TFS-Assets/GSD/posters/adme_genes_axiom_assays_ashg_2019_poster.pdf.
7. Bycroft C, Freeman C, Petkova D et al. (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562:203–209.


Learn more at **thermofisher.com/microarrays**

applied biosystems