Detection of Copy Number Variations in Blood Typing Research and ADME Genes Using Axiom Assays

Ram Varma, Kay Aull, Carsten Bruckner, Jeremy Gollub, Alan Roter, Jeanette Schmidt.
Thermo Fisher Scientific, Santa Clara, California 95051, USA.

ABSTRACT

Copy number variations (CNVs) in *CYP2D6* and other genes that are relevant for drug absorption, distribution, metabolism, and excretion (ADME) are difficult to detect because these genes are members of highly homologous gene families, or because of the high prevalence of deletions and duplications in the population. Genotyping blood type variants is complicated by the frequent deletion of the *RHD* gene and its high homology with *RHCE*, which together encode for Rh factor antigens. Copy number–aware genotyping of markers in these regions significantly improves the accuracy of variant calling of genes with high homology.

Results from experiments run on the Applied Biosystems™ Axiom™ PMD Array are presented here. The 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) *de novo* whole-genome 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, copy number states are determined by implementation of a Hidden Markov Model in single-sample mode. Breakpoints are discovered and copy number segments are assigned a state. Fixed-region analysis uses an optimized multisample clustering algorithm to assign copy number states to each region in each sample.

CNVs in *RHD*, *CYP2D6*, *CYP2A6*, *GSTT1*, *GSTM1*, and *UGT2B17* were identified during replicate runs of more than 90 HapMap samples. Discovery analysis detected CNVs with overall analytical reproducibility greater than 90% for losses and greater than 70% for gains. It was less sensitive in the detection of duplications in a small 150 bp region around exon 9 of *CYP2D6*, which is important in predicting the ability of individuals to metabolize drugs. However, using fixed-region CNV analysis, duplications and deletions were detected in 100% concordance with known truth in this region in all samples and were reproducible in all but one sample. Across all defined fixed regions of interest, homozygous deletions were detected in 100% concordance with known truth, while single-copy losses were detected with 100% analytical sensitivity and over 95% analytical specificity. In summary, the platform enables analysis with both methods and displays superior analytical sensitivity and specificity for known small regions while enabling discovery for larger regions and across the whole genome.

INTRODUCTION

The Axiom PMD Array is a comprehensive imputation-aware genotyping resource with more than 850,000 markers designed to drive deeper scientific insights into complex disease susceptibility, pharmacogenomics, and genetic factors underlying ancestry and lifestyle in diverse populations. It provides accurate genotyping of predictive markers in genes, including *CYP2D6* and *CYP2A6,* that are in highly homologous regions with known relevance to drug metabolism, and genes such as *RHD* and *RHCE* that are important for blood typing and transfusion medicine research.

The Axiom PMD Array is enabled for copy number analysis in targeted regions as well as for *de novo* genome-wide copy number discovery [1]. The targeted copy number regions of interest on this array are shown in Table 1. Multiple probe sets (typically around 50) were selected based on sequence homology and other considerations [2], and assigned to the genes or regions of interest at the time of array design. Probe sets used for copy number analysis may interrogate nonpolymorphic or SNP markers in the region. To obtain accurate copy number estimates, dense tiling of probe sets was needed in the shorter regions.

Gene/region of interest	Approximate length (bp)
RHD	58,050
GSTM1	10,800
UGT2B17	34,100
CYP2A6	2,700
CYP2A6 (intron 2–intron 4)	2,500
CYP2A6 (5′ flank)	4,650
CYP2D6 (3′ flank)	450
CYP2D6 (exon 9)	150
CYP2D6 (exon 1–exon 8)	4,050
CYP2D6 (5′ flank)	3,850
GSTT1	8,800

Table 1. Targeted copy number regions of interest on the Axiom PMD Array.

Estimates of copy number in the ADME genes enable copy number–aware genotyping and the production of star-allele translation tables to assist in generating reports to discover ADME properties of drugs. Detecting events in the exon 9 region of *CYP2D6* is especially important in identifying gene conversion events and *CYP2D6–2D7* hybrid genes [3]. Copy number calls on four regions spanning *CYP2D6* in HapMap samples run on two 96-format Axiom PMD Arrays are shown in Figure 1. Several samples show changes in copy number state in the exon 9 region. CNV analysis also enables the use of an algorithmic approach to more accurately genotype the *RHD* gene [4]. This is illustrated in Figure 2, which shows the genotype calls plot for a probe set interrogating an *RHD* marker.

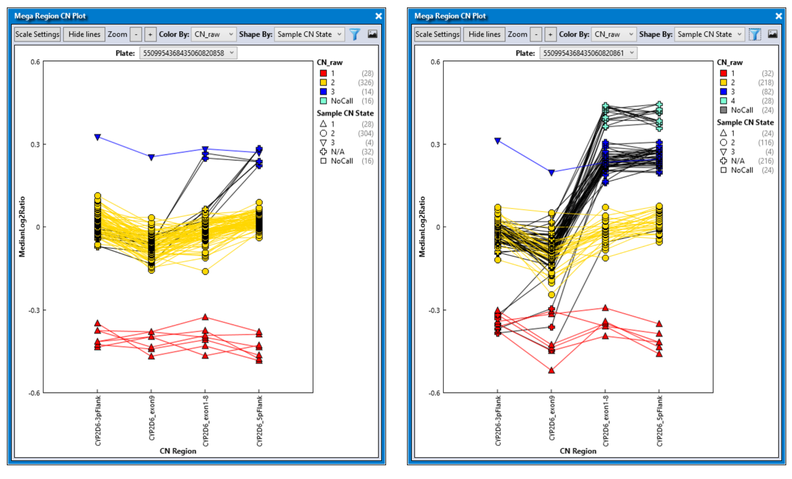

Figure 1. Detection of CNV events in *CYP2D6*. Copy number calls on four *CYP2D6* regions from two Axiom PMD Array plates that were used to analyze HapMap CEU or CHB/JPT samples. Changes in copy number state are shown along the gene in many samples. Plots were generated using Applied Biosystems™ Axiom™ Analysis Suite.

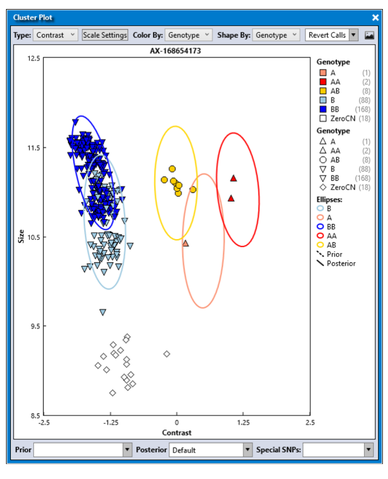

Figure 2. Genotyping of an *RHD* variant. Genotype calls plot for a probe set interrogating NG_007494.1(RHD):g.17121G>T variant in HapMap CEU, CHB/JPT, and YRI samples run on three 96-format Axiom PMD Arrays. Using *RHD* copy number, the algorithm accurately detects all genotypes of interest [4].

MATERIALS AND METHODS

96 unique samples from Coriell 1000 Genomes and HapMap collections were prepared using standard protocols and assayed on the Axiom PMD Array at our R&D laboratory in Santa Clara, California. Replicate samples were run on multiple 96-format array plates. Copy number analysis was performed using the two methods—fixed-region analysis and whole-genome discovery—available in Applied Biosystems™ Array Power Tools (APT) v2.10.2 and Axiom Analysis Suite v4.0.1 software packages, and sample quality filtering was performed in accordance with Best Practices Workflows and established guidelines [5]. Copy number calls were compared to Applied Biosystems™ TaqMan® Assays, internal data from other arrays, or published data including expected frequencies.

RESULTS

Results shown here are from 93 unique samples from the Coriell collections—73 samples with 4 replicates, 19 samples with 3 replicates, and 1 sample with 2 replicates—that passed all quality metrics. Figures 3 and 4 show $\log_2$ ratio tracks for replicates of two samples.

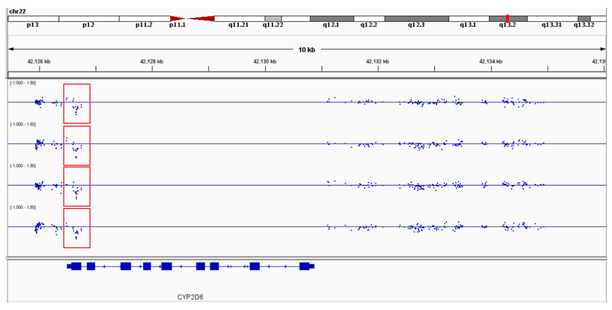

Figure 3. $\log_2$ ratio tracks for 4 replicates of a sample assayed on the Axiom PMD Array showing a single-copy deletion on *CYP2D6* exon 9. $\log_2$ ratio of total intensity for a probe set is calculated with respect to a reference intensity representing normal diploid state at the corresponding marker. Values close to the zero line indicate diploid status. Losses are characterized by negative values. For this sample, probe sets in the exon 9 region show a negative dip in $\log_2$ ratio, indicating a local copy number loss. Analysis was performed using Axiom Analysis Suite, and results were viewed in Integrative Genomics Viewer.

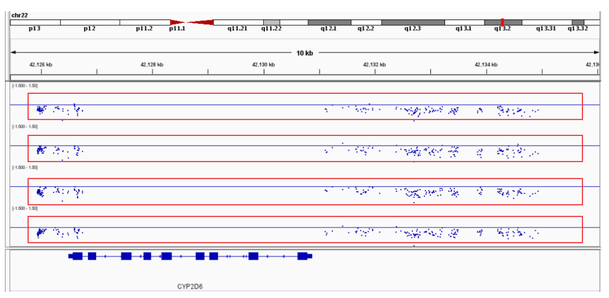

Figure 4. $\log_2$ ratio tracks for 4 replicates of a sample assayed on the Axiom PMD Array showing a single-copy deletion spanning the *CYP2D6* gene and the flanking regions. For this sample, all probe sets in the region have negative $\log_2$ ratios indicating single-copy loss of the whole gene.

Overall performance was evaluated using concordance of copy number calls with known truth or across replicate samples (Table 2). Using fixed-region analysis, homozygous deletions were detected in 100% concordance with known truth, single-copy losses were detected with 100% analytical sensitivity and over 95% analytical specificity, and single-copy gains were detected with over 95% analytical sensitivity and specificity.

CN state	Performance evaluation for 93 samples and 9 copy number regions of interest
CN0	236/236 concordant replicate calls over 64 unique events in all samples and regions.
CN1	617/617 concordant replicate calls over 161 unique events in all samples and regions.
CN2	2,147/2,149 concordant replicate calls in all samples and regions. Discordant (CN1 or CN3) calls in 2 samples.
CN3	153/157 concordant replicate calls over 50 unique events in all samples and regions. Discordant (CN2) calls in 3 samples.

Table 2. Results of concordance of calls on 9 copy number (CN) regions from a total of 351 samples—93 unique samples with replicates—assayed on the Axiom PMD Array.

Copy number calls on *CYP2D6* exon 9 were fully concordant, except for one sample where a single-copy gain was not called consistently (Table 3).

CN state	Performance evaluation on *CYP2D6* exon 9
CN1	39/39 concordant replicate calls over 10 unique events.
CN2	288/288 concordant replicate calls.
CN3	22/24 concordant replicate calls over 6 unique events.

Table 3. Results of concordance of calls on *CYP2D6* exon 9 from a total of 351 samples—93 unique samples with replicates—assayed on the Axiom PMD Array.

Using *de novo* discovery analysis, deletions and duplications that overlapped genes or regions of interest were concordant more than 90% and 70% of the time, respectively. On *CYP2D6* exon 9, copy gains were called in 17 of 77 normal diploid samples (false positives) while copy gains were not called in 3 of 6 samples with known events (false negatives). Fixed-region analysis was clearly superior in detecting events in this 150 bp region.

CONCLUSIONS

The Axiom platform enables copy number analysis with two methods:

- The fixed-region method results in superior analytical sensitivity and specificity for known small regions
- The *de novo* discovery method enables detection of events for larger regions and across the whole genome

The Axiom PMD Array enables copy number calling in several important ADME and blood typing research genes with analytical sensitivities and specificities exceeding 95%.

REFERENCES

1. Thermo Fisher Scientific. Technical note: Axiom copy number analysis. Pub. No. COL32811.
2. Aull K, Bruckner C, Varma R et al. (2019) Homology-aware design facilitates microarray CNV detection within gene families. Blood transfusion Genomics Consortium. PgmNr 492/W. Poster presented at ASHG 2019.
3. Gaedigk A, Ingelman-Sundberg M, Miller NA et al. (2018) The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. *Clinical Pharmacology and Therapeutics* 103:399–401.
4. Nelson CS, Black D, Aull D et al. (2019) Genotyping in regions of high homology. Blood transfusion Genomics Consortium. PgmNr 2483/F. Poster presented at ASHG 2019.
5. Thermo Fisher Scientific. Axiom Analysis Suite v4.0.1 user guide.

ACKNOWLEDGEMENTS

The authors would like to thank members of the microarray R&D laboratory at Thermo Fisher Scientific, Santa Clara, CA, for their help in designing and performing the experiments on which this study is based.

For Research Use Only. Not for use in diagnostic procedures.