Copy Number Variation Analysis on the HapMap DNA Samples Using TaqMan® Gene Dosage Assays for CYP2D6, CYP2E1, CYP2A6, GSTM1 and GSTT1

Applied Biosystems

Kelly Li[1], Catalin Barbacioru[1], Fiona Hyland[1], Timothy T. Harkins[1], Kashif A. Haque[2], Robert A. Welch[2] and Katherine Lazaruk[1],   [1] MCB Division, Applied Biosystems, Foster City, CA and [2]Core Genotyping Facility, Division of Cancer Epidemiology and Genetics, SAIC Frederick, National Cancer Institute, Gaithersburg, MD

ABSTRACT

Gene copy number variation is now recognized as an important type of polymorphism in the human genome. Accurate detection of copy number changes is critical for understanding how gene copy number variation plays a role in disease or drug response.   Here we report our development of an application using real-time PCR assays to quantify gene copy number and the results of copy number for 5 important drug metabolism genes (CYP2D6, CYP2E1, CYP2A6, GSTM1 and GSTT1) using 270 DNA samples from the International HapMap Project. Our copy number analysis reveals that copy number variation for these 5 genes with the HAPMAP sample panel and significant copy number frequency differences of these genes in different population. In addition, these gene copy number data is also useful in explaining the failure of SNP assay and identifying the inaccuracy of genotyping data due to departure from diploidy. For example, our copy number analysis showed high percentage of individuals from the HapMap panel had the null allele(s) for these 5 genes for GSTM1 and GSTT1 genes.   These observations suggest why many of the SNPs in GSTM1 and GSTT1 SNPs failed QC (were qc-) as reported by the HapMap Project, as they failed due to a low pass rate <80%, a low Hardy-Weinberg p-value, or to Mendelian inconsistencies. All of these inconsistencies within the SNP genotyping data can be explained when the gene dosage data are factored into the analysis.

INTRODUCTION

Gene duplication/deletion can have a significant impact on phenotype. Copy number polymorphisms have been associated with genetic diseases such as cancer, immunological and neurological disorders, and also associated with variations in drug response as previously demonstrated. We have developed TaqMan gene dosage assays for 5 important DME genes. Measuring copy number variation in these genes is an important complement to genotyping for two reasons. First, falsely assuming diploidy can lead to inaccuracy in determining the genotype. Second, copy number can interact with genotype to influence phenotype of a variant. The availability of robust and validated genotyping and copy number variation assays allows for full characterization and understanding of genotypes of the genes of interest.

MATERIALS AND METHODS

Primers and probes were designed on genomic DNA sequence of by an AB proprietary Assay Design Pipeline. The 270 individual samples from International HAPMAP Project were used (Table1). 15 samples were spotted as duplicates. Four replicate plates were run for each assay.   Each well is duplexed with two assays. The FAM™ dye-based assay is designed to detect the genes-of-interest and the VIC® dye-based assay is for the reference gene, RNase P. The RNase P gene is known to have 2 copies per diploid genome regardless of the status of the gene-of-interest.   The final assay condition is 10ng of gDNA, 900 nM of each primer and 250nM of FAM-MGB probe for the gene-of-interest, 100nM of each primer and 250nM VIC-TAMRA™ probe for RNase P in 1X TaqMan® Universal Master Mix in a 20 μl assay volume. The assays are performed on the Applied Biosystems 7900HT Sequence Detection System and the thermal-cycling conditions are: 2 mins at 50°C, 10 mins at 95°C, followed by 40 cycles of 15 secs at 92°C and 60 secs at 60°C. Real-time data is collected by the SDS 2.2 software. Relative quantitation is used to determine copy number and associated confidence by a newly developed algorithm, which identifies outlier measurements from all wells and uses linear models to estimate gene copy number for each sample, and provides also confidence levels in predicted copy numbers.

RESULTS

Table1. HAPMAP Sample Summary

Population	Sample Sources	Sample Types	Total Samples
Caucasian	European ancestry from CEFH	30 Trios	90
African	Yoruba of Ibadan, Nigeria	30 Trios	90
Chinese	Beijing	45 individuals	45
Japanese	Tokyo	45 individuals	45
total			270

Figure 1. Evaluation of Sample Duplicate Consistency
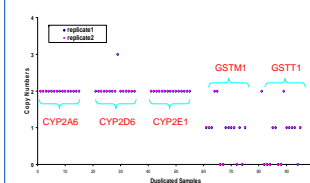
Fifteen samples, covering all 4 populations, from the 270 unique individuals are spotted twice on each plate. Sample duplicate consistency of copy number call is evaluated for each of the 5 assays. Perfect consistency of sample duplicates is observed in all the population and copy numbers (0, 1, 2, and 3).

Figure 2. Copy Number from Trios Follows Mendelian Inheritance

YRI family #48 is taken as a demonstration to show copy number transmission in trio samples for all 5 genes. Mendelian Inheritance is followed.
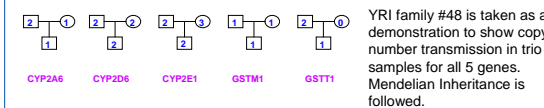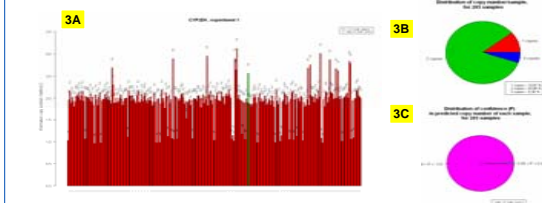
Figure 3. Copy Number Determination of CYP2D6 Gene

CYP2D6 is used as an example of copy number determination for the HAPMAP panel. Fig. 2A Copy number is determined for individual samples and copy numbers of one, two and three are predicted. Fig.2B shows the copy number distribution for all the samples. Majority of individuals have 2 copies. Around 10% and 5% samples have one or three copies respectively. Fig. 2C reveals distribution of confidence for predicted copy number for each sample. For all the samples, the confidence level is above 0.95 and furthermore, 99% samples have confidence between 0.99 and 1.

### Figure 4. Copy Number Frequencies in Each Population
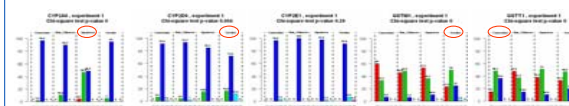
Copy number frequencies in all four populations for all five genes are calculated. All five genes except CYP2E1 show significant differences in copy number frequencies among the population by Chi-square test. Red circles indicate the population that is different from others.

Table 2. Pair-wised Comparison for Copy Number Frequencies between the Populations

	CYP2A6c1	CYP2D6c1	CYP2E1c1	GSTT1c1	GSTM1c1
hh_AfAm_Eur	0	0	0	1	1
hh_AfAm_Chn	0	0	0	0	1
hh_AfAm_Jpn	1	0	0	0	1
hh_Eur_Chn	0	0	0	1	0
hh_Eur_Jpn	1	0	0	1	0
hh_Chn_Jpn	1	0	0	0	0

Significance of difference in copy number frequencies between populations is tested. One in red indicates that there is a significant difference between populations (P < 0.05); Zero indicates that there is no significant different between populations. There is a significant difference between the Japanese (for CYP2A6), Caucasian (for GSTM1), African (for GSTT1) from all other populations.

Figure 5. Copy Number Analysis Is An Important Complement to SNP Genotyping for GSTM1 (Fig. 5A) And GSTT1 (Fig. 5B).
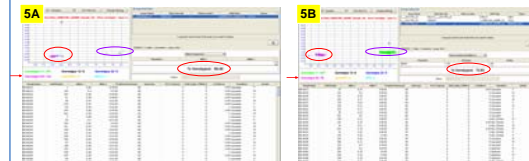
Many of the SNP assays in GSTM1 and GSTT1 failed QC (were qc-) as reported by the HapMap project as well as in DME assays. Our copy number analysis reveals high percentage of individuals from the HapMap panel had the null allele(s) for GSTM1 and GSTT1 genes. Fig. 5A shows this GSTM1 assay has only 56% pass rate. Among the 124 samples that are not genotypable, 123 individuals have null alleles(zero copy number). Among the 161 samples that are called homozygous, 112 of them (69.6%) have one copy of GSTM1. Fig. 5B shows the GSTT1 assay has only 73% pass rate. Among the 78 failed individuals, 77 of them have null alleles (zero copy). Among the 207 samples that are called homozygous, 132 of them (63.9%) have one copy of GSTT1. Therefore, departures from diploidy cause genotyping failure and furthermore homozygous allele by genotyping may only have one copy, which could have impact on phenotype. Thus, measuring copy number variation in these genes is an important complement to genotyping.

CONCLUSIONS

We have developed an application using real-time PCR assays to quantify gene copy number for 5 important DME genes. In this study we determine copy numbers of these five genes for 270 individuals from HAPMAP sample panel. The copy number analysis for these genes show perfect consistency for sample duplicates, high Hardy-Weinberg p-values (data not shown) and Mendelian inheritance.   Copy number variation (from 0 to 3) is observed for all 5 genes. Significant copy number frequency difference in different population is revealed for CYP2A6, CYP2D6, GSTM1 and GSTT1. We demonstrate departures from diploidy can cause apparent genotyping failure and give inaccurate genotyping for GSTM1 and GSTT1. Therefore, measuring copy number variation in these genes is an important complement to genotyping assays. The availability of robust and validated genotyping and copy number variation assays allows for full characterization and understanding of genotypes of the genes of interest.