Population-specific \textit{GSTM1} copy number variation

R. Stephanie Huang\textsuperscript{1}, Peixian Chen\textsuperscript{2}, Steve Wisel\textsuperscript{1}, Shiwei Duan\textsuperscript{1}, Wei Zhang\textsuperscript{1}, Edwin H. Cook\textsuperscript{4}, Soma Das\textsuperscript{2}, Nancy J. Cox\textsuperscript{3} and M. Eileen Dolan\textsuperscript{1,*}

\textsuperscript{1}Section of Hematology-Oncology, Department of Medicine, \textsuperscript{2}Department of Human Genetics and \textsuperscript{3}Section of Genetic Medicine, Department of Medicine, The University of Chicago, 5841 S. Maryland Ave., PB Box MC2115, Chicago, IL 60637, USA and \textsuperscript{4}Department of Psychiatry, University of Illinois at Chicago, Chicago, IL 60612, USA

Received August 10, 2008; Revised and Accepted October 15, 2008

As one of the major glutathione conjugation enzymes, GSTM1 detoxifies a number of drugs and xenobiotics. Its expression and activity have been shown to correlate both with cancer risks and drug resistance. Through a genome-wide association study, we identified a significant association between HapMap SNP rs366631 and \textit{GSTM1} expression. In this study, utilizing lymphoblastoid cell lines derived from International HapMap Consortium CEU and YRI populations, we designed and performed site-specific genotyping assays for both rs366631 and a highly homologous \textit{GSTM1} upstream site. Copy number variation (CNV) assays were performed for three different regions of the \textit{GSTM1} gene. We demonstrated that HapMap SNP rs366631 is a non-polymorphic site. The false genotyping call arises from sequence homology, a common \textit{GSTM1} region deletion and a non-specific genotyping platform used to identify the SNP. However, the HapMap call for rs366631 genotype is an indicator of \textit{GSTM1} upstream region deletion. Furthermore, this upstream deletion can be used as a marker of \textit{GSTM1} gene deletion. Using a novel \textit{GSTM1} CNV assay, we showed a population-specific CNV in this region upstream of the gene. More than 75% of the Caucasian (CEU) samples exhibit \textit{GSTM1} deletion and none contain two copies of \textit{GSTM1}. In contrast, up to 25% of African (YRI) samples were found to have two copies of \textit{GSTM1}. In conclusion, HapMap rs366631 is a pseudo-SNP that can be used as a \textit{GSTM1} deletion marker. Both the pseudo-SNP allele frequency and \textit{GSTM1} upstream region CNV show population-specific patterns between CEU and YRI samples.

\textbf{INTRODUCTION}

Glutathione \textit{S}-transferase (GST), a major Phase II family of conjugation enzymes, plays a crucial role in the detoxification of drugs and environmental carcinogens (1). In humans, cytosolic GSTs are divided into eight classes: \(\alpha\) (GSTA), \(\mu\) (GSTM), \(\pi\) (GSTP), \(\theta\) (GSTT), \(\tau\) (GSTZ), \(\sigma\) (GSTS), \(\omega\) (GSTO) and \(\kappa\) (GSKT), each of which contains one or more of the homodimeric or heterodimeric isoforms (2). To date, a total of five \(\mu\) class genes (\textit{M1}–\textit{M5}) situated in tandem on chromosome 1p13.3 have been identified (3). \textit{GSTM1}, a member of GSTM family enzymes, detoxifies a number of electrophilic substances, including carcinogens such as polycyclic aromatic hydrocarbons, ethylene oxide, epoxides and styrene (3). Lack of \textit{GSTM1} enzyme may impair metabolic elimination of carcinogenic compounds from the body, thereby increase cancer risk (4), while overexpression of \textit{GSTM1} isozyme has been shown to associate with chemotherapeutic resistance (5,6). Interestingly, our laboratory recently found a significant inter-ethnic difference in \textit{GSTM1} expression in a set of 176 HapMap lymphoblastoid cell lines (LCLs) derived from individuals of European or African descent (7) (Fig. 1).

Given the functional importance of \textit{GSTM1} in cellular protection from environmental and oxidative stress, genetic variations that alter this gene expression or enzyme activity may play an important role in both risk of disease development and cellular sensitivity to drugs. Two polymorphisms of the \textit{GSTM1} gene, namely \textit{GSTM1*0} and \textit{GSTM1*1/GSTM1*0}, have been identified. \textit{GSTM1*0} is a deleted allele, and the homozygous allele (\textit{GSTM1-null genotype}) expresses no \textit{GSTM1} protein. Most studies of \textit{GSTM1} polymorphism have compared the homozygous deletion genotype with the genotypes containing at least one functional allele. Common
-expression. The total number of samples is plotted against GSTM1 expression obtained through Affymetrix GeneChip® Human Exon 1.0 ST array. Open bars represent samples with HapMap rs366631 TT genotype call; grey bars represent samples with HapMap rs366631 CT genotype call and solid black bars represent samples with HapMap rs366631 CC genotype call.

homozygous gene deletions have been reported for GSTM1 with variable frequency in different ethnic groups (4,8). Alleles GSTM1A and GSTM1B differ by a C > G substitution at base position 534. This C > G substitution results in the substitution of Asn > Lys at amino acid 172 (9); however, this substitution results in no functional difference between the two alleles (10). As a result, GSTM1A and GSTM1B are categorized together as non-null conjugator phenotypes. Numerous studies have shown the positive correlation between GSTM1*0 (homozygous deletion) and increased risk of cancers (11–15) and with better treatment outcomes (16–18).

Utilizing a set of EBV-transformed LCLs from the International HapMap Consortium, our laboratory has recently performed a genome-wide association study between over 2 million HapMap SNPs and global gene expression (19). In an effort to identify relationships between genetic variants and gene expression for a set of important pharmacogenes in humans (7), we identified a previously unknown significant association between a SNP (rs366631) genotype and GSTM1 expression in cell lines derived from individuals with either Caucasian (CEU) or African (YRI) ancestry. rs366631 is located 11 265 bp downstream of GSTM1 gene on chromosome 1 with a population-specific minor allele frequency (0.18 and 0.49 for CEU and YRI, respectively). Further examination revealed that the TT genotype of rs366631 is associated with little or no GSTM1 expression, and the C allele (minor allele) carriers are associated with higher GSTM1 expression in both CEU and YRI samples (Fig. 1). The objective of our current study is to validate this predictor of GSTM1 expression.

RESULTS

Genotyping

UCSC Genome Brower shows two duplications of more than 1000 bases of non-repeat masked sequence flanking the GSTM1 gene region (Fig. 2). rs366631 is located in one of these duplicons downstream of GSTM1 (http://genome.ucsc.edu). In contrast to a T being the common allele for rs366631, at the corresponding site in the upstream duplication, a C allele is the common allele. Due to the high homology between these two duplicons, we designed and performed two sequence-specific genotyping assays, which included sequence-specific PCR, single-base extension (SBE) reaction followed by denaturing high-performance liquid chromatography (DHPLC). PCR products from both the rs366631 site and the upstream C site were sequenced and results confirmed that the amplified sequences were specific to the region of interest. We genotyped 30 randomly selected HapMap samples for both the rs366631 site and the upstream C site, and all samples genotyped showed non-polymorphic TT genotype at the rs366631 site. Approximately half of the samples tested showed non-polymorphic amplification of the C allele at the upstream C site and the remainder showed no amplification at the upstream C site (Table 1). The sequencing results were consistent with the SBE-DHPLC results.

To insure non-polymorphic TT genotyping calls were not due to the sample selection, we also tested 52 non-HapMap CEPH (Centre d’Etude du Polymorphisme Humain) and an additional 28 HapMap samples that had HapMap rs366631 CC calls, all samples showed TT genotype at the rs366631 site, further suggesting HapMap rs366631 is non-polymorphic.

Upstream C region quantification

To evaluate the relative abundance of the genomic region that includes the upstream C site, we performed quantitative PCR (qPCR) assays on 176 HapMap CEU and YRI LCLs. We found a significant association between HapMap rs366631 genotyping calls and the relative amount of C region genomic sequence (\(P = 2 \times 10^{-79}\), when association was analyzed with all 176 samples using population and sex as covariates). The CC genotype was found to correlate with two copies, CT with one copy and TT with zero copies of the GSTM1 upstream C site region (Fig. 3). Furthermore, the relative amount of the C region genomic sequence was significantly correlated with GSTM1 expression (measured by Affymetrix GeneChip® Human Exon 1.0 ST array, \(P < 0.0001\)). Upon evaluating the ‘pseudo genotype’ and C region copy number relationship, we identified an ethnic-specific GSTM1 upstream C region copy number variation (CNV), for which ~75% of CEU samples showed a complete deletion of the upstream C region, 25% had a single copy and none of the CEU samples contained two copies of C region DNA (Fig. 3). In contrast, 25% of the YRI samples showed deletion, 50% had a single copy and 25% had two copies of the C region. The deletion allele frequency is 0.86 and 0.51 for CEU and YRI, respectively. This was consistent with Hardy–Weinberg equilibrium (\(P > 0.05\)) and showed a significant population difference (Fst = 0.15).
GSTM1 region CNV

This upstream C region has been reported to be within GSTM1 gene deletion by Kerb et al. (20). To study whether the relative abundance of upstream C region quantity is part of the GSTM1 gene deletion, we performed two additional GSTM1 CNV assays and compared results to those obtained from our upstream C region assay (Table 2). A total of 14 samples were randomly selected from all three rs366631 genotyping groups. The upstream C region CNV results are in 100% agreement with those obtained through Applied Biosystems (ABI) TaqMan GSTM1 copy number assay (evaluating exon 1 region) and through an in-house GSTM1 copy number assay (evaluating intron 6/exon 7 region). Given the different genomic regions these assays have evaluated, our findings support that the upstream C region is a component of the GSTM1 gene deletion.

DISCUSSION

In this study, we evaluated in greater depth the relationship between the HapMap reported SNP, rs366631, and GSTM1 gene expression. Our site-specific SBE-DHPLC assay designed to genotype rs366631, demonstrated that rs366631 is a non-polymorphic site. Given the high homology between the duplicon where rs366631 is located and the duplicon upstream of GSTM1 gene, it is possible that HapMap genotyping calls were evaluating both sites. We therefore designed an additional site-specific SBE-DHPLC assay for the upstream C site. All amplified C site PCR products showed a CC genotype; however, ~50% were not amplified. When comparing our site-specific genotyping results with HapMap rs366631 calls, we reasoned that non-amplification of the C site was consistent with a TT call for rs366631; while a CT call for rs366631 was likely resulting from the presence of one copy of the upstream C site and a CC call for rs366631 corresponding to two copies.

The platform used to make the genotype calls for HapMap rs366631 was Illumina HapMap-BeadArrays. The Illumina HapMap-BeadArray utilized GoldenGate Assay design, which contains a set of three oligonucleotides. Two of them are 5’ assay oligos designed to hybridize upstream of the SNP, with the 3’ base of each being allele-specific. The third oligo is located 3’ end of the two allele-specific oligos and
Table 1. Genotyping results for randomly selected 30 HapMap samples at the rs366631 site and the upstream C site as compared with HapMap rs366631 calls

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Sample id</th>
<th>C Site SBE assay</th>
<th>T/T SBE assay</th>
<th>Sample id</th>
<th>C Site SBE assay</th>
<th>T/T SBE assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>YRI</td>
<td>NA19200</td>
<td>C/C</td>
<td>T/T</td>
<td>NA19201</td>
<td>C/C</td>
<td>T/T</td>
</tr>
<tr>
<td>CEU</td>
<td>NA128291</td>
<td>C/T</td>
<td>T/T</td>
<td>NA19202</td>
<td>C/T</td>
<td>T/T</td>
</tr>
<tr>
<td>CEU</td>
<td>NA11849</td>
<td>C/T</td>
<td>T/T</td>
<td>NA19201</td>
<td>C/T</td>
<td>T/T</td>
</tr>
<tr>
<td>YRI</td>
<td>NA18502</td>
<td>C/T</td>
<td>T/T</td>
<td>NA18870</td>
<td>C/T</td>
<td>T/T</td>
</tr>
<tr>
<td>YRI</td>
<td>NA19201</td>
<td>C/T</td>
<td>T/T</td>
<td>NA19137</td>
<td>C/T</td>
<td>T/T</td>
</tr>
<tr>
<td>YRI</td>
<td>NA19144</td>
<td>C/T</td>
<td>T/T</td>
<td>NA19144</td>
<td>C/T</td>
<td>T/T</td>
</tr>
<tr>
<td>YRI</td>
<td>NA19203</td>
<td>C/T</td>
<td>T/T</td>
<td>NA19143</td>
<td>C/T</td>
<td>T/T</td>
</tr>
<tr>
<td>YRI</td>
<td>NA19143</td>
<td>C/T</td>
<td>T/T</td>
<td>NA18545</td>
<td>C/T</td>
<td>T/T</td>
</tr>
<tr>
<td>CHB</td>
<td>NA18545</td>
<td>C/T</td>
<td>T/T</td>
<td>NA18504</td>
<td>C/T</td>
<td>T/T</td>
</tr>
<tr>
<td>CHB</td>
<td>NA18566</td>
<td>C/T</td>
<td>T/T</td>
<td>NA18842</td>
<td>C/T</td>
<td>T/T</td>
</tr>
<tr>
<td>JPT</td>
<td>NA18546</td>
<td>C/T</td>
<td>T/T</td>
<td>NA06985</td>
<td>C/T</td>
<td>T/T</td>
</tr>
<tr>
<td>CEU</td>
<td>NA11995</td>
<td>C/T</td>
<td>T/T</td>
<td>NA12981</td>
<td>T/T</td>
<td>T/T</td>
</tr>
<tr>
<td>CEU</td>
<td>NA12981</td>
<td>T/T</td>
<td>T/T</td>
<td>NA11993</td>
<td>T/T</td>
<td>T/T</td>
</tr>
<tr>
<td>CEU</td>
<td>NA11993</td>
<td>T/T</td>
<td>T/T</td>
<td>NA18504</td>
<td>T/T</td>
<td>T/T</td>
</tr>
<tr>
<td>CEU</td>
<td>NA11984</td>
<td>T/T</td>
<td>T/T</td>
<td>NA18526</td>
<td>T/T</td>
<td>T/T</td>
</tr>
<tr>
<td>CEU</td>
<td>NA11881</td>
<td>T/T</td>
<td>T/T</td>
<td>NA18586</td>
<td>T/T</td>
<td>T/T</td>
</tr>
<tr>
<td>CEU</td>
<td>NA11881</td>
<td>T/T</td>
<td>T/T</td>
<td>NA18945</td>
<td>T/T</td>
<td>T/T</td>
</tr>
<tr>
<td>YRI</td>
<td>NA19153</td>
<td>T/T</td>
<td>T/T</td>
<td>NA18945</td>
<td>T/T</td>
<td>T/T</td>
</tr>
<tr>
<td>YRI</td>
<td>NA18546</td>
<td>T/T</td>
<td>T/T</td>
<td>NA18967</td>
<td>T/T</td>
<td>T/T</td>
</tr>
<tr>
<td>CHB</td>
<td>NA18526</td>
<td>T/T</td>
<td>T/T</td>
<td>NA18986</td>
<td>T/T</td>
<td>T/T</td>
</tr>
<tr>
<td>JPT</td>
<td>NA18964</td>
<td>T/T</td>
<td>T/T</td>
<td>NA18967</td>
<td>T/T</td>
<td>T/T</td>
</tr>
</tbody>
</table>

Figure 3. Relationship between HapMap rs366631 genotype and relative amounts of upstream C region in both CEU and YRI samples. Circles represent CEU samples; diamonds represent YRI samples. P-values for association between genotype and upstream CNV are $P = 2 \times 10^{-32}$ and $P = 2 \times 10^{-37}$ for CEU and YRI, respectively.

is locus-specific (http://www.ncbi.nlm.nih.gov.proxy.uchicago.edu/SNP/snp_viewTable.cgi?method_id=915). Given the high homology for two duplicons flanking GSTM1 region, this design may non-specifically produce signals from both duplicons. Therefore, SNP calls are made when, in fact, two non-polymorphic homologous sites are evaluated using a single assay. Furthermore, literature supports homologous GSTM1 deletion frequencies of 20–80% in various ethnic populations (4,8). Considering both the genotyping platform design and the potential inclusion of the upstream C site in the common GSTM1 deletion, it became clear that when the GSTM1 region is deleted, a TT call is made; however, we were puzzled that CC calls were made when both C and T sites were present. One possibility is that the Illumina HapMap BeadArray oligos were designed to capture the upstream C site nucleotide and as a consequence also nonspecifically captured the downstream T site nucleotide at a decreased efficiency. After quantifying the relative prevalence of the upstream C region, a significant association was found between HapMap rs366631 genotyping calls and the relative amount of C site genomic sequence. When two copies of C site genomic regions exist, a CC call is made. When only one copy of C site region is present, a CT call is made. When there was a homozygous deletion of this region, TT genotype was called. These findings support the possibility that the relative abundance of the upstream C region is responsible for HapMap genotype calls. However, other possible explanation may also result in the wrong genotyping calls (e.g. genotype calling algorithm forces the call of the three possible genotypes). Without an evaluation of the raw genotype intensity data that is not currently in the public domain, we can only speculate the cause. Interestingly, rs366631 is listed as a marker on both Illumina Human 1M-duo (http://www.illumina.com/pages.ilmn?ID=261&source=home) and Human 610-quad beadchips (http://www.illumina.com/pages.ilmn?ID=248). However, the genotype results on this site are read as a ‘no call’ (M. Kibriya, H. Ahsan, C. Brown and K. White, personal communication). In addition, the rs366631 SNP genotype is not currently reported in the Pha-seIII HapMap data release (http://www.hapmap.org/cgi-perl/gbrowse/hapmap3_B36/).

In addition, we found a significant correlation between GSTM1 genomic deletion (measured by qPCR using primers designed to the upstream C region) and GSTM1 expression. This was consistent with previous findings by French et al. that GSTM1-null genotype corresponds to lower GSTM1 expression when compared with non-null genotype (21). Furthermore, McLellan et al. have shown that duplication of GSTM1 gene may cause ultra rapid GSTM1 enzyme activity (22). Interestingly, in our study, all CEU HapMap samples exhibit at least partial deletion at the GSTM1 upstream C region; while ~25% of YRI samples have two copies of the
C region. This ethnic-specific genomic deletion is a likely explanation for the observed population difference in \(\text{GSTM1}\) expression between the \(\text{CEU}\) and \(\text{YRI}\) samples. \(\text{GSTM1}\) deletion polymorphism has generally been described as either null (homozygous deletion) or non-null. Although significant inter-ethnic differences in \(\text{GSTM1}\) homozygous deletion have been reported (4,8), our study is the first to show ethnic differences in the number of expressed \(\text{GSTM1}\) alleles between samples with Caucasian ancestry and those with African ancestry, which corresponds to differences in \(\text{GSTM1}\) expression. A considerable functional difference between carriers of two versus one active \(\text{GSTM1}\) allele has been suggested (20). Given the important role \(\text{GSTM1}\) in disease risk and drug response, the functional significance of this ethnic-specific \(\text{GSTM1}\) genetic CNV should be further explored. Finally, although the low \(\text{GSTM1}\) allele frequency in \(\text{CEU}\) samples may justify studies using null versus non-null \(\text{GSTM1}\) genotype analysis, when studying the genetic contribution of \(\text{GSTM1}\) in populations that have fairly high \(\text{GSTM1}\) allele frequency such as the Yorubans, a more quantitative analysis of \(\text{GSTM1}\) allele is warranted.

Two additional \(\text{GSTM1}\) CNV assays were performed and compared with the upstream \(\text{C}\) region CNV assay. We found complete agreement between these assays that interrogated different sections of the \(\text{GSTM1}\) genomic region in randomly selected 14 HapMap LCL samples. When two copies of the upstream region are present, this encompasses two copies of exon 1 and intron 6/exon 7 DNA; but if there is no copies of the upstream region, exon 1 and intron 6/exon 7 DNA are not present. Upon evaluating the Database of Genomic Variants (http://projects.tcag.ca/variation/ accessed on 10 July 2008; Human genome build 36), nine patterns of \(\text{GSTM1}\) region deletions are currently listed. Three of these patterns contain large deleted regions encompassing both the T (\(\text{GSTM1}\) downstream) and C (\(\text{GSTM1}\) upstream) sites (23–25); while the other four are smaller deletion regions which include neither \(\text{T}\) nor \(\text{C}\) site (22,26–28). In addition, one deletion pattern included only the \(\text{T}\) site (23). The final pattern included only the \(\text{C}\) site in the deletion region (23), which is in agreement with our findings. In addition, although not listed in the Database of Genomic Variants, Kerb \textit{et al.} have reported an approximately 18 500 bp deletion, which includes the upstream \(\text{C}\) region, but not the downstream \(\text{T}\) region (20). When combining three CNV assays, our data suggest the potential deletion region extends at least from the upstream \(\text{C}\) region through the exon 7 region of \(\text{GSTM1}\), a region longer than that reported by Kidd \textit{et al.} (23) and within the deletion region described by Kerb \textit{et al.} (20). These data suggest either potential individual specific CNV patterns due to sample selection (Kidd \textit{et al.} studied eight individuals and we evaluated 14 individuals with one sample that overlapped between the two studies), or a potential new pattern of \(\text{GSTM1}\) region CNV that warrants further study.

In summary, through site-specific genotyping, we demonstrated that HapMap SNP rs366631 is a non-polymorphic site. The sequence homology, the common \(\text{GSTM1}\) region deletion, and non-specific genotyping platform usage are likely contributing to the false genotyping call. Although referred to as HapMap rs366631, this genotype is an indicator of \(\text{GSTM1}\) upstream region deletion. Furthermore, this upstream deletion can be used as a marker of \(\text{GSTM1}\) gene deletion. Lastly, we developed a new \(\text{GSTM1}\) CNV assay that interrogates upstream of \(\text{GSTM1}\) and demonstrates a population-specific CNV in this region.

## MATERIALS AND METHODS

### DNA isolation

EBV-transformed B-LCLs were purchased from the Coriell Institute for Medical Research (Camden, NJ). Cell lines were maintained and diluted as described previously (29). DNA was isolated from 87 HapMap CEU, 89 YRI, 8 CHB (Chinese from Beijing), 8 JPT (Japanese from Tokyo) LCLs and 52 non-HapMap CEPH LCLs. For each cell line, a total of \(5 \times 10^6\) cells were pelleted, washed in ice-cold PBS and centrifuged to remove PBS. All cell pellets were flash-frozen and stored at \(-80^\circ\text{C}\) until DNA isolation. Total DNA was extracted using the 5Prime kit (5Prime Inc., Fisher Scientific, Pittsburgh, PA) following the manufacturer’s protocol. DNA quality assessment and quantification were conducted using the optical spectrometry 260/280 nm ratio.

### Genotyping

For specificity, two region-specific PCR amplification primer sets were designed with each one specifically amplifying one duplinuc to produce 151 bp PCR products (Table 3). Both the rs366631 site and upstream \(\text{C}\) site genotyping were performed on a set of randomly selected 30 HapMap samples which consist of 10 CEU, 10 YRI, 5 CHB and 5 JPT. In addition, rs366631 genotyping was performed on 52 non-HapMap CEPH samples and 28 HapMap samples (22 YRI, 3 CHB and 3 JPT) that had CC genotyping calls. Sample IDs can be found in the Supplementary Material.

### Table 3. PCR and SBE primer design for rs366631, upstream C site and an endogenous control

<table>
<thead>
<tr>
<th>Region of interest</th>
<th>PCR primer sets</th>
<th>SBE primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs366631 (\text{GSTM1}) downstream</td>
<td>F: 5’-TGG CCT CAC AGT GAA TTC TT-3’&lt;br&gt;R: 5’-TCT TTC ACA GCT CAA GAC CC-3’</td>
<td>5’-CAT TAT GGA TTA AGA GTT TAT GAT T-3’&lt;br&gt;none</td>
</tr>
<tr>
<td>(\text{GSTM1}) upstream C site</td>
<td>F: 5’-TGG CCT CAC AGT GAA CTC TT-3’&lt;br&gt;R: 5’-TCT TTC ACA GCT CAA GAC GG-3’</td>
<td>5’-CAT TAT GGA TTA AGA GTT TAT GAT T-3’&lt;br&gt;none</td>
</tr>
<tr>
<td>huB2M</td>
<td>F: 5’-TGGCACCCTGCTGAGATGACTG-3’&lt;br&gt;R: 5’-GGATGCTAGGACACGAGGAC-3’</td>
<td>none</td>
</tr>
</tbody>
</table>

SBE, single-base pair extension. Bold represent nucleotides specific for the region of interest.
PCR reactions were set up in a total volume of 15 \mu l containing 2.5 mM MgCl₂, 50 \mu M each dNTP, 125 nM forward and reverse primers, 0.625 U AmpliTaq Gold (Applied Biosystems, Foster City, CA, USA) and 30 ng genomic DNA. Reactions were denatured initially at 95°C for 15 min followed by 40 cycles of 95°C for 15 s, 58°C for 15 s and 72°C for 30 s and a final extension at 72°C for 10 min. PCR products were cleaned-up with Shrimp Alkaline Phosphatase and Exonuclease I at 37°C for 45 min. To confirm region-specific amplification, PCR products were sequenced as described later.

Genotyping was conducted through SBE reactions followed by DHPLC. A volume of 7.2 \mu l of purified PCR product was added for SBE reaction with 1 \mu M SBE primer, 250 \mu M each ddNTP and 1.5 U Thermo Sequenase (GE Healthcare) in a 12 \mu l volume. SBE reactions were run under the following conditions: 96°C for 2 min, followed by 60 cycles at 96°C for 30 s, 55°C for 30 s and 60°C for 30 s. SBE products were denatured and separated on a WAVE 3500HT DHPLC System, and analyzed by Navigator software (Transgenomic, Inc.).

Sequencing
The sequencing reactions were carried out in a total volume of 10 \mu l containing 0.5 or 1 \mu l purified PCR product, 400 nM forward or reverse primer, 2 \mu l Big Dye terminator v3.1 (Applied Biosystems) and 1 \times Big Dye dilution buffer. The reactions were denatured initially at 95°C for 5 min then cycled at 95°C for 10 s, 50°C for 5 s and 60°C for 4 min for 25 cycles of the standard sequencing conditions. Same T or C site amplification primer sets were used for both forward and reverse direction sequencing. Sequencing was performed on an ABI 3100 Sequencer (Applied Biosystems). The sequence chromatograms were analyzed using the Sequencher 4.5 program.

Upstream C region quantification
Prevalence of the C site genomic region was quantified in 176 HapMap CEU and YRI LCLs. We performed SYBR green qPCR assays using the same primer sets designed for upstream C site amplification described earlier and a primer set for an endogenous control (huB2M) (Table 3) on the Applied Biosystems 7500 real-time PCR system. Reactions were carried out in 25 \mu l volumes consisting of 12.5 \mu l 2 \times SYBR Green PCR Master mix (Applied Biosystems), 1.5 \mu l primer mix (final of 300 nM for each forward and reverse primers) and 40 ng DNA. The thermocycler parameters were: 50°C for 2 min, 95°C for 10 min and 40 cycles of 95°C for 15 s/60°C for 1 min for ABI assay; 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 s/61.5°C for 1 min for the in-house assay.

SUPPLEMENTARY MATERIAL
Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS
The authors are grateful for excellent technical support provided by Shannon M. Delaney and Wasim K. Bleibel.

Conflict of Interest statement. Dr Dolan has presented seminars at meetings funded by Affymetrix, Inc.

FUNDING
This Pharmacogenetics of Anticancer Agents Research (PAAR) Group (http://pharmacogenetics.org) study was supported by NIH/NIGMS grant U01GM61393 and U01GM61374 (http://pharmgkb.org/). This study was also supported by P50 CA125183 University of Chicago Breast Cancer SPORE grant. R.S.H. receives support from NIH grant 5 T32 GM 007019-30.
REFERENCES


