Multi-ethnic cytochrome-P450 copy number profiling: novel pharmacogenetic alleles and mechanism of copy number variation formation

S Martis¹, H Mei¹, R Vijzelaar², L Edelmann¹, RJ Desnick¹ and SA Scott¹

To determine the role of CYP450 copy number variation (CNV) beyond CYP2D6, 11 CYP450 genes were interrogated by multiplex ligation-dependent probe amplification and quantitative PCR in 542 African-American, Asian, Caucasian, Hispanic and Ashkenazi Jewish individuals. The CYP2A6, CYP2B6 and CYP2E1 combined deletion/duplication allele frequencies ranged from 2 to 10% in these populations. High-resolution microarray-based comparative genomic hybridization (aCGH) localized CYP2A6, CYP2B6 and CYP2E1 breakpoints to directly oriented low-copy repeats. Sequencing localized the CYP2B6 breakpoint to a 529-bp intron 4 region with high homology to CYP2B7P1, resulting in the CYP2B6*29 partial deletion allele and the reciprocal, and novel, CYP2B6/2B7P1 duplicated fusion allele (CYP2B6*30). Together, these data identified novel CYP450 CNV alleles (CYP2B6*30 and CYP2E1*1C2X) and indicate that common CYP450 CNV formation is likely mediated by non-allelic homologous recombination resulting in both full gene and gene-fusion copy number imbalances. Detection of these CNVs should be considered when interrogating these genes for pharmacogenetic drug selection and dosing.

Keywords: CYP2A6; CYP2B6; CYP2E1; copy number variation; pharmacogenetics; pharmacogenomics

INTRODUCTION

The significance of structural variation in human disease and phenotypic diversity has increasingly become recognized and several studies have generated catalogs of copy number variants (CNVs) in the human genome to facilitate a better understanding of their functional relevance. It is estimated that ~12% of the human genome contains CNVs, which range in size from ~1 kb to several megabases. A number of mechanisms of CNV formation have been proposed including non-allelic homologous recombination, NHER, non-homologous end joining, fork stalling and template switching, and microhomology-mediated break-induced replication. Like single-nucleotide polymorphisms, CNVs not only influence disease susceptibility but also interindividual differences in drug response, as several important pharmacogenetic genes are known to have variable copy number (for example, CYP2D6, GSTT1, GSTM1, SULT1A1 and UGT2B17). The polymorphic cytochrome-P450 (CYP450) system is a superfamily of 56 functional genes (and over 50 highly homologous pseudogenes) that are principally involved in drug metabolism and bioactivation. Inherited genetic variation among the CYP450 genes contributes to disease susceptibility and interindividual differences in drug response. While single-nucleotide polymorphisms represent an important form of CYP450 variation that can influence enzyme activity, CNVs also can affect CYP450 activity by altering gene expression. However, most studies on CYP450 CNVs have focused on CYP2D6 and to a lesser extent on CYP2A6, with limited CNV population data on other CYP450 family members. This is primarily because CNV detection is challenging for some CYP450 genes due to the presence of homologous gene family members, pseudogenes and complex rearrangements including chimeras and gene conversions. For example, CYP2D6 CNV characterization and metabolizer phenotype assignment is convoluted by the presence of CYP2D6/2D7P1 hybrid alleles and the fact that not all duplication alleles encode functional enzymes (for example, increased function CYP2D6*2xN, reduced function CYP2D6*10xN and non-functional CYP2D6*4xN).

Thus, this study aimed to determine the prevalence of CNVs in CYP450 genes other than CYP2D6 in a large multi-ethnic cohort of healthy, unrelated African-American, Asian, Caucasian, Hispanic and Ashkenazi Jewish (AJ) individuals. Copy number was determined for 11 CYP450 genes by multiplex ligation-dependent probe amplification (MLPA) and validated by quantitative PCR (qPCR) copy number assays. Additionally, a subset of samples were subjected to high-resolution microarray-based comparative genomic hybridization (aCGH) to accurately map the size of identified CNVs and localize breakpoint intervals. In addition to revealing the diverse frequency profiles of CYP450 CNVs across multi-ethnic populations, these studies also identified two novel CYP450 alleles, CYP2B6*30 and CYP2E1*1C2X. Moreover, the frequency data offer insight into the adaptive evolution of CYP450 CNVs and the breakpoint analyses provide insight into their mechanism of formation.

MATERIALS AND METHODS

Study population

Peripheral blood samples from healthy donors who indicated their racial/ethnic background and gave informed consent for the use of their DNA for research were obtained from the New York Blood Center with Institutional Review Board approval.
Multi-ethnic CYP450 copy number variation
S Martis et al

All aCGH experiments were performed according to manufacturer’s instructions and as previously described.33–34 In brief, 1 μg each of experimental and reference DNAs were digitalized with AluI and Rsal restriction endonucleases (Promega, Madison, WI, USA) and fluorescently labeled with Cy3anine 5-dCTP (Cy-5; experimental) and Cy5-anine 3-dCTP (Cy-3; reference) using the Genome DNA Labeling kit (Agilent Technologies). Labeling, hybridization and reference DNAs were purified, denatured, pre-annealed with Cot-1 DNA (Invitrogen, Carlsbad, CA, USA) and blocking reagent (Agilent Technologies), and hybridized to arrays in a rotating oven (20 h.p.m.) at 65 °C for 24 h. After hybridization and recommended washes, the arrays were scanned at 5 μm resolution with a G2505C Agilent Microarray Scanner. Images were processed and the data analyzed with CytoGenomics 1.5 software (both from Agilent Technologies). All array data passed the recommended quality control metrics. Breakpoints were defined as interprobe regions with surrounding average log2 ratios that indicated a change in copy number from two to one or three copies.

CYP2B6 and CYP2E1 sequence analyses

The breakpoint regions (~3 kb) of the CYP2B6 and CYP2E1 fusion alleles (CYP2B6*29 and *30) were assessed by breakpoint-spanning PCR using overlapping primers (Supplementary Table S3) and the SequaPrep Long PCR Kit (Invitrogen). In addition, long-range PCR was performed to generate CYP2B6 and CYP2E1 gene fusion-specific amplicons of 18 and 11 kb spanning exons 1–5 and exons 4–9, respectively, which were used as templates for exon and exon-intron junction sequencing using primers listed in Supplementary Table S3. DNA samples from four controls with two copies of CYP2B6 by MLPA and qPCR were included as negative controls. Long-range PCRs were performed in 20 μl containing ~100 ng of DNA, SequaPrep 1× PCR buffer, 0.4 μl DMSO, SequaPrep 1× Enhancer A, 0.5 μM of forward and reverse primers (Supplementary Table S3), and 1.8 units of SequaPrep Long Polymerase. Amplification consisted of an initial denaturation step at 94 °C for 2 min followed by 10 amplification cycles (94 °C for 10 s, 58 °C for 30 s and 68 °C for 1 min per kb), 30 amplification cycles (94 °C for 10 s, 58 °C for 30 s and 68 °C for 1 min per kb (~20 s per cycle) and a final extension at 72 °C for 5 min. Additionally, for selected subjects, all CYP2B6 exons and exon/intron boundaries were individually amplified and sequenced using exon primers listed in Supplementary Table S3.

For subjects harboring a CYP2E1 duplication, long-range PCR was performed to generate a CYP2E1-specific amplicon of ~3 kb spanning the promoter region and exon 1 using conditions detailed above and primers listed in Supplementary Table S4. Exons 2–9 of CYP2E1 were also amplified and directly sequenced using primers listed in Supplementary Table S4.

All amplicons were digested with 3.0 units of shrimp alkaline phosphatase and 2.0 units of Exonuclease I (both from USB Corporation, Cleveland, OH, USA) and bidirectionally sequenced using amplification and/or sequencing primers. Sequencing was performed on an ABI 3700 Sequencer (Applied Biosystems) and sequence chromatograms were analyzed using Mutation Surveyor software v3.30 (SoftGenetics).

RESULTS

CYP450 copy number profiling

CNVs were interrogated with an MLPA panel of 11 CYP450 genes (CYP1A1, CYP1A2, CYP1B1, CYP2A6, CYP2B6, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4 and CYP3A5) among 542 healthy, unrelated African-American, Asian, Caucasian, Hispanic and AJ individuals. In addition to CYP2D6 (data not shown), CNVs were detected in CYP2A6, CYP2C9 and CYP2E1. Representative MLPA data from samples with variable CYP450 copy number are shown in Table 1 and Supplementary Table S1, and the identified CYP2A6, CYP2B6 and CYP2E1 allele frequencies are summarized in Table 2. Wide variability in CYP2A6, CYP2B6 and CYP2E1 deletion/duplication allele frequencies was observed across the tested populations ranging from 0 to 8.3% for deletions and 0 to 4.1% for duplications. The total CNV allele frequencies were 9.0% for African-Americans, 10.3% for Asians, 4.4% for Caucasians, 5.5% for Hispanics and 2.4% for the AJ (Table 2). All detected CYP2A6, CYP2B6 and CYP2E1 CNVs were validated by TaqMan qPCR copy number assays, indicating that the multiplexed MLPA assay was robust at detecting CYP450 CNVs. The genomic locations of MLPA and TaqMan probes are illustrated in Supplementary Figure S1.
Breakpoint identification by aCGH

The CYP2A6, CYP2B6 and CYP2E1 breakpoint intervals were further refined by aCGH using a custom high-density oligonucleotide microarray on a subset of 46 samples with CYP450 CNVs as determined by MLPA and qPCR (Table 3). The University of California Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/; human genome build GRCh37/hg19) was used to visualize the location of known human genes from the NCBI reference sequence collection (RefSeq), structural variants from the Database of Genomic Variants (DGV; http://projects.tcag.ca/dgv/main/) and low-complexity repetitive elements (for example, short interspersed elements (SINEs) and long interspersed elements (LINEs)).

CYP2A6

Both CYP2A6 deletion and duplication alleles were detected by MLPA with copy number probes in exons 1, 2 and 5. The deletion allele was detected in all populations except the AJ and was most common among Asians (8.3%). In contrast, the duplication allele frequency was highest among the AJ (2.0%) but not detected in the Caucasian or Hispanic populations. Two TaqMan qPCR assays (exons 1 and 7; Supplementary Figure S1) confirmed the MLPA results, the downstream TaqMan probe indicated a normal balanced copy number, suggesting the possibility of partial CYP2A6 deletion and duplication (Figure 2).

High-resolution aCGH mapped the size of the CYP2B6 deletion and duplication to ~32 kb at 19q13.2 (~41.33–41.36 Mb) and localized the CNV breakpoints to a cluster of Alu and LINE elements within the 3’ flanking regions of CYP2A6 and the more distal homologous CYP2A7 gene (Figure 1). The identified breakpoints were mapped to two directly oriented LCRs: chr19:41 336 557–41 361 522 (proximal) and chr19:41 367 421–41 393 307 (distal) (Table 3). Sequence alignment indicated that the two LCRs share >90% sequence homology and encompass both CYP2A6 and CYP2A7 (Figure 1).

CYP2B6

Identification of the CYP2B6/2B7P1 duplicated fusion allele (CYP2B6*30). Both CYP2B6 deletion and duplication alleles were detected in the African-American and Asian populations, respectively, by MLPA with copy number probes in exons 2, 3 and 4. Two TaqMan qPCR assays were used to confirm these results, one in exon 4 and a second located 12.6 kb downstream of the CYP2B6 gene (Supplementary Figure S1). No commercial TaqMan probes were available that specifically interrogated exons 5–9 of CYP2B6. Interestingly, whereas the exon 4 TaqMan probe confirmed the MLPA results, the downstream TaqMan probe indicated a normal balanced copy number, suggesting the possibility of partial CYP2B6 deletion and duplication (Figure 2).

High-resolution aCGH mapped the size of the CYP2B6 deletion and duplication to ~70 kb at 19q13.2 (~41.45–41.51 Mb) and localized the CNV breakpoints to a homologous region within intron 4 of both CYP2B6 and the proximal CYP2B7P1 pseudogene (Figure 2). The identified breakpoints mapped to two directly oriented LCRs: chr19:41 426 446–41 460 038 (proximal) and chr19:41 494 916–41 528 965 (distal) (Table 3). Sequence alignment indicated that the two LCRs share >90% sequence homology (Figure 2). These results suggested that the partial deletion allele represented the previously reported CYP2B6*29 allele36 and the partial duplication represented a reciprocal, and novel, CYP2B6/2B7P1 duplicated fusion allele (Figures 2 and 3). The data for this allele, including the sequencing described below, were reviewed by the Cytochrome P450 Allele Nomenclature Committee37,38 and it subsequently has been designated as CYP2B6*30 (http://www.cypalleles.ki.se/cyp2b6.htm).

Sequence analyses of CYP2B6*29 and *30. A long-range PCR-based strategy was employed to interrogate the breakpoint regions of the identified CYP2B6*29 deletion and CYP2B6*30

Table 1. Representative MLPA data from selected samples with variable CYP450 copy number

| Gene   | MLPA probe | Samples (peak ratios) |
|--------|------------|-----------------------|
|        |            | AA404 | AA410 | HISP287 | C048 | C106 |
| CYP2A6 | Exon 1     | 0.983 | 1.540 | 0.924   | 0.521 | 0.974 |
|        | Exon 2a    | 1.029 | 1.487 | 0.953   | 0.575 | 1.038 |
|        | Exon 2b    | 1.036 | 1.430 | 0.934   | 0.580 | 0.966 |
|        | Exon 5     | 1.001 | 1.386 | 1.057   | 0.549 | 1.033 |
| CYP2B6 | Exon 2     | 0.576 | 1.035 | 0.963   | 1.060 | 1.008 |
|        | Exon 3     | 0.563 | 1.111 | 0.928   | 1.022 | 0.963 |
|        | Exon 4     | 0.535 | 1.030 | 0.950   | 1.022 | 1.003 |
|        | Exon 5     | 0.943 | 0.916 | 1.521   | 0.899 | 1.412 |
|        | Exon 6     | 0.999 | 1.016 | 1.519   | 0.954 | 1.406 |
|        | Exon 8     | 1.024 | 0.926 | 1.514   | 0.998 | 1.466 |

Table 2. Variable CYP450 copy number allele frequencies

| Gene   | Copies/allele | African-American (n = 210) | Asian (n = 204) | Caucasian (n = 206) | Hispanic (n = 218) | Ashkenazi Jewish (n = 246) |
|--------|---------------|-----------------------------|----------------|---------------------|------------------|--------------------------|
|        |               | Frequency | 95% CI   | Frequency | 95% CI   | Frequency | 95% CI   | Frequency | 95% CI   | Frequency | 95% CI   |
| CYP2A6 | 2             | 0.010     | 0.000–0.023 | 0.015     | 0.000–0.032 | 0.015     | 0.000–0.032 | 0.020     | 0.000–0.038 |
|        | 1             | 0.952     | 0.923–0.981 | 0.902     | 0.861–0.943 | 0.985     | 0.969–1.000 | 0.986     | 0.971–1.000 | 0.980     | 0.962–1.000 |
|        | 0             | 0.038     | 0.012–0.064 | 0.083     | 0.045–0.121 | 0.015     | 0.000–0.031 | 0.014     | 0.000–0.029 | 0.014     | 0.000–0.029 |
| CYP2B6 | 2             | 0.005     | 0.000–0.014 | —         | —         | —         | —         | —         | —         | 0.000     | 0.000–0.005 |
|        | 1             | 0.995     | 0.986–1.000 | 0.995     | 0.985–1.000 | 1.000     | 1.000–1.000 | 1.000     | 1.000–1.000 | 1.000     | 1.000–1.000 |
|        | 0             | 0.005     | 0.000–0.014 | —         | —         | —         | —         | —         | —         | 0.000     | 0.000–0.005 |
| CYP2E1 | 2             | 0.038     | 0.012–0.064 | —         | —         | —         | —         | —         | —         | 0.000     | 0.000–0.002 |
|        | 1             | 0.962     | 0.936–0.988 | 1.000     | 1.000–1.000 | 0.971     | 0.948–0.994 | 0.959     | 0.933–0.985 | 0.996     | 0.988–1.000 |
|        | 0             | —         | —         | —         | —         | —         | —         | —         | —         | —         | —         |

Abbreviations: AA, African-American; C, Caucasian; HISP, Hispanic; MLPA, multiplex ligation-dependent probe amplification.

*See ‘Materials and methods’ for peak ratio calculation. Shading indicates MLPA probes with variable copy number (one copy >0.25 and <0.75; three copies >1.25 and <1.7; four copies >1.7 and <2.0).
Table 3. Summary of CYP450 aCGH analyses

| Gene   | Chr | CNV interval (Mb) | CNV size (kb) | Proximal LCR | Distal LCR | Repetitive elements                          |
|--------|-----|-------------------|---------------|--------------|------------|---------------------------------------------|
| CYP2A6 | 19  | 41.33–41.36       | ~32           | 41 336 557–41 361 522 | 41 367 421–41 393 307 | SINEs and LINEs in directly oriented LCRs; CYP2A7 homologous gene |
| CYP2A7 | 19  | 41.45–41.51       | ~70           | 41 426 446–41 460 038 | 41 494 916–41 528 965 | SINEs in directly oriented LCRs; CYP2B7P1 pseudogene |
| CYP2E1 | 10  | 135.250–135.381   | ~130          | 135 236 151–135 249 933 | 135 380 989–135 394 059 | SINEs in directly oriented LCRs; SPRN and SPRNP1 pseudogene |

Abbreviations: aCGH, microarray-based comparative genomic hybridization; CNV, copy number variant; kb, kilobases; LCR, low-copy repeat (segmental duplication); LINEs, long interspersed elements; Mb, megabases; SINEs, short interspersed elements (e.g., Alu).

Based on human genome build GRCh37/hg19.

Figure 1. CYP2A6 copy number confirmation and breakpoint region identification. (a) TaqMan qPCR copy number assays located in CYP2A6 exons 1 and 7 were used to confirm CNVs detected by MLPA. Note the homozygous duplications and deletions for AJ 080 and Asian 035, respectively. (b) Representative CYP2A6 aCGH data, indicating that the CNV breakpoints localized to two segmental duplications surrounding the CYP2A6 gene, resulting in full gene deletion and duplication. Red, black and blue dots represent oligonucleotide aCGH probes with log2 ratios less than −0.20, in between −0.20 and 0.20, and >0.20, respectively. (c) Illustration of the local genomic architecture surrounding CYP2A6 with enlarged views of the interval illustrated in (b) (~41.32–41.40 Mb). The location of known human genes from NCBI RefSeq, aCGH probes, identified CNV interval, segmental duplications, repetitive elements (for example, SINEs, LINEs and LTRs), and structural variants from the DGV (blue = copy number gain; red = copy number loss; brown = copy number gain and loss) are represented. The CNV breakpoints map to a cluster of neighboring repetitive Alu and LINE elements present in two directly oriented LCRs (>90% similarity). (d, e) Dot Matrix plots showing regions of similarity based upon alignment using BLAST (http://blast.ncbi.nlm.nih.gov). Both the x and y axes represent nucleotide sequence (in kilobases) corresponding to (d) the genomic interval displayed in (b) and (c), and (e) the proximal and distal segmental duplications highlighted by red boxes in (c). Alignments are shown in the plots as lines, whereby directly oriented homologies are slanted from the bottom left to the upper right corner. Approximate location of CYP2A6 CNV region is highlighted in (d) with a red box.
duplication fusion alleles (Figure 3). All breakpoint spanning amplicons were sequenced and the putative breakpoint regions were identified by multiple sequence alignment using CLUSTAL (http://www.ebi.ac.uk/clustalw). Based on the alignment data, the breakpoint region was localized to a 529-bp interval within introns 4 and exons 5 of both CYP2B6 and CYP2B7P1 where the sequences of the two genes are nearly identical. Breakpoint junctions for CYP2B6*29 and *30 were defined at the locations where the
nucleotide sequence changed from CYP2B7P1 to CYP2B6 and from CYP2B6 to CYP2B7P1, respectively (Figure 3; Supplementary Figures S2 and S3). The homologous introns 4 of CYP2B6 and CYP2B7P1 also contain four directly oriented Alu elements (AluSc, AluSx, AluSx and AluSg/x).

In addition, all CYP2B6*29 and *30 exons and exon-intron junctions were sequenced. The two alleles were translated using ExPasy (http://web.expasy.org) and aligned with CYP2B6 to identify polypeptide sequence alterations. Despite the unique breakpoint region identified in our study, the coding sequence of our CYP2B6*29 allele was consistent with that previously reported36 (http://www.cypalleles.ki.se/cyp2b6.htm), sharing complete sequence identity with exons 1–4 of CYP2B7P1 and exons 5–9 of CYP2B6. Conversely, CYP2B6*30 shared complete sequence identity with exons 1–4 of CYP2B6 and exons 6–9 of CYP2B7P1. Given the high sequence homology between exons 5 of CYP2B6 and CYP2B7P1, it is unclear which gene this exon is derived from for CYP2B6*30. Fifteen amino-acid alterations were detected in CYP2B6*30 (p.A279P, p.N289I, p.L290I, p.R326K, p.I332V, p.E339A, p.H341D, p.Y354R, p.S360A, p.R378X, p.I381T, p.H397R, p.T423N, p.M459V and p.R487C), indicating that the duplicated CYP2B6*30 allele is most likely non-functional. Sequencing the CYP2B6 exons not included in *30 detected the p.Q172H and p.K262R variants; however, the phase of these alleles could not be determined so as to assess the breakpoint region identified in our study, the coding sequence of our CYP2E1*1Cx2 allele was consistent with that previously reported36 (http://www.cypalleles.ki.se/cyp2e1.htm), sharing complete sequence identity with exons 1–4 of CYP2E1*1Dx2 and exons 5–9 of CYP2E1. Conversely, CYP2E1*30 shared complete sequence identity with exons 1–4 of CYP2E1 and exons 6–9 of CYP2E1*1Dx2. Moreover, these data indicate that common CYP450 CNV formation is likely mediated by NAHR resulting in both full gene and gene-fusion copy number imbalances.

CYP2E1

**Identification of the duplicated CYP2E1 allele.** Duplicated CYP2E1 alleles were detected by MLPA with probes in exons 5, 6 and 7 in all tested populations except for Asians, and were most frequent among Hispanics (4.1%). No CYP2E1 deletion carriers were identified. Six TaqMan qPCR assays targeting exons 1 and 9, and two each upstream (45.0 and 16.8 kb) and downstream (13.8 and 15.2 kb) of CYP2E1 were used to confirm the MLPA data (Supplementary Figure S1). All qPCR copy number data were consistent with those obtained by MLPA and included full gene duplication (Figure 4).

High-resolution aCGH mapped the mean size of the CYP2E1 duplication to ~130 kb at 10q26.3 (~135.250–135.381 Mb) and localized the CNV breakpoint to two directly oriented LCRs: chr10: 135 236 151–135 249 933 (proximal) and chr10: 135 380 989–135 394 059 (distal) (Table 3). Sequence alignment indicated that the two LCRs share >98% sequence homology and encompass the majority of the SPRN gene and its distal SPRNP1 pseudogene (Figure 4).

**Sequence analyses of the CYP2E1*1Cx2 duplication.** Although the size of the CYP2E1 duplication (~130 kb) prevented the design of a duplication allele-specific PCR, common CYP2E1 variants were interrogated by direct sequencing of all samples harboring a CYP2E1 duplication. No CYP2E1*1Dx2, *3, *4, *5, *6 or *7 variants were identified among the duplication carriers; however, two 5′ promoter region repeat alleles were identified, *4C (6 repeats) and *1D (8 repeats).39 For heterozygous *1C/*1D carriers (n = 7; 30% of duplication carriers), the phase could not be determined so the duplicated allele for these cases would either be *1Cx2 or *1Dx2. However, for homozygous *1C/*1C subjects (n = 16; 70% of duplication carriers), the duplicated allele must also be *1C. For these subjects, all nine CYP2E1 exons were sequenced and no additional coding region variants were identified. These data also have been reviewed by the Cytochrome P450 Allele Nomenclature Committee and the duplication allele has been designated as CYP2E1*1Cx2 (http://www.cypalleles.ki.se/cyp2e1.htm). Notably, no CYP2E1*1Dx homzygotes were detected among the CYP2E1 duplication subjects, suggesting that the CYP2E1*1Dx2 allele is either uncommon or non-existent in the tested populations.

**DISCUSSION**

The paucity of CNV data among the CYP450 genes, with the notable exception of CYP2D6, prompted our copy number survey of CYP1A1, CYP1A2, CYP1B1, CYP2A6, CYP2B6, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4 and CYP3A5 in the African-American, Asian, Caucasian, Hispanic and AJ populations. Of the 11 genes analyzed, CNVs were detected and confirmed in CYP2A6, CYP2B6 and CYP2E1, with multi-ethnic frequencies ranging from 0 to 8.3% for deletions and 0 to 4.1% for duplications. Interestingly, the discordant CYP450 duplication and deletion allele frequencies (for example, CYP2A6 deletion (8.3%) and duplication (1.5%) frequencies among Asians) suggest that selective pressures likely have acted on certain alleles. CYP2D6 CNVs were also detected in all studied populations; however, these data are being integrated with CYP2D6 genotyping data for an independent manuscript in preparation. In addition to identifying the CNV frequencies for this panel of CYP450 genes, high-resolution aCGH and sequencing analyses facilitated the identification of the novel CYP2B6*30 duplicated fusion allele and CYP2E1*1Cx2. Moreover, these data indicate that common CYP450 CNV formation is likely mediated by NAHR resulting in both full gene and gene-fusion copy number imbalances.

CYP2A6 is a hepatic CYP450 enzyme primarily involved in the metabolism of nicotine, cotinine and nitrosamine precarcinogens.40 The CYP2A6 gene is highly polymorphic and variant alleles have been associated with interindividual variability in nicotine metabolism, smoking behavior and tobacco-related cancer risk.41 We detected wide variability in the distribution of CYP2A6 deletion alleles across the five tested populations with the highest frequency observed among Asians (8.3%). The duplication allele was detected at low frequencies (<2%) among the African-Americans, Asians and AJ, and was not detected in either the Caucasian or Hispanic cohorts. Although deletions are often under stronger purifying selection pressure than duplications to remove deleterious variants from the population,42 the high CYP2A6 deletion frequency observed in our Asian population could be due to its previous association with lower cancer risks among Asians43 and, therefore, may actually confer a protective effect. Moreover, it has been suggested that the CYP2A6 deletion is a relatively recent genetic event that likely occurred after the Caucasoid/Asian split ~35 000 years ago.44

Using high-resolution aCGH, we mapped the size of the CYP2A6 deletion and duplication to ~32 kb, with breakpoints localizing to the 3′ regions of CYP2A6 and the distal homologous CYP2A7 gene. Both CYP2A6 and CYP2A7 are encompassed by two directly oriented LCRs with >90% sequence homology and the identified breakpoints clustered to repetitive Alu and LINE elements. NAHR during meiosis can generate rearrangements as a consequence of recombination between regions of high sequence homology like LCRs and dispersed Alu repetitive elements.45 For CYP2A6, the presence of two direct LCRs likely renders the region susceptible to rearrangements, leading to NAHR-mediated reciprocal full gene deletion and duplication.22 As such, the different subtypes of the CYP2A6 deletion allele (CYP2A6*4A-H) previously reported in African populations46 are likely the result of unique breakpoints mediated by different homologous Alu and/or other neighboring low-complexity repeats. Moreover, gene conversions of the CYP2A6 3′ flanking region by CYP2A7,47 which would not be detected by the methods employed in the current study, are also likely mediated by these low-complexity repeats.

CYP2B6 is involved in the metabolism of anticancer agents (for example, cyclophosphamide and ifosfamide), anesthetics (for example, ketamine and propofol), and antitretrovirals (for example,
EFV and NVP. Commonly tested variant CYP2B6 alleles include p.K262R (*4), p.R487C (*5), and p.[Q172H*K262R] (*6). Notably, the CYP2B6*6, *18 and *26 decreased activity alleles have been shown to influence the pharmacokinetics of the non-nucleoside reverse transcriptase inhibitor efavirenz, which may have utility for CYP2B6 genotype-guided dose adjustment. However, little is known about CNVs at this locus other than the CYP2B6*29 partial deletion allele.

By MLPA screening and qPCR assays, we identified both CYP2B6 deletion and duplication alleles that were limited to exons 1–4 of the CYP2B6 gene. High-resolution aCGH mapped the size of the CYP2B6 deletion/duplication to ~70 kb and localized the CNV breakpoints to introns 4/exons 5 of CYP2B6 and the proximal CYP2B7P1 pseudogene, both within LCRs that share >90% sequence homology. Similarly to CYP2A6, these LCRs likely mediate NAHR between CYP2B6 and CYP2B7P1, resulting in the previously reported CYP2B6*29 partial deletion allele and the reciprocal, and novel, CYP2B6/2B7P1 duplicated fusion allele (designated as CYP2B6*30 by the Cytochrome P450 Nomenclature Committee). The CYP2B6*29 and *30 breakpoints were further

**Figure 4.** CYP2E1 copy number confirmation and breakpoint region identification. (a) TaqMan qPCR copy number assays located in exons 1 and 9 and surrounding CYP2E1 were used to confirm CNVs detected by MLPA. (b) Representative CYP2E1 aCGH data, indicating that the CNV breakpoints localized to two segmental duplications surrounding CYP2E1, resulting in full gene duplication. Red, black and blue dots represent oligonucleotide aCGH probes with log2 ratios less than –0.20, in between –0.20 and 0.20, and >0.20, respectively. (c) Illustration of the local genomic architecture surrounding CYP2E1 with enlarged views of the interval illustrated in (b) (~135.21–135.41 Mb). Genome browser tracks are as detailed in Figure 1. The CNV breakpoints map to highly homologous directly oriented LCRs (~98% similarity). (d, e) Dot Matrix plots showing regions of similarity based upon alignment using BLAST (http://blast.ncbi.nlm.nih.gov). Both the x and y axes represent nucleotide sequence (in kilobases) corresponding to (d) the genomic interval displayed in (b) and (c), and (e) the proximal and distal segmental duplications highlighted by red boxes in (c). Alignments are shown in the plots as lines, whereby directly oriented homologies are slanted from the bottom left to the upper right corner. Approximate location of CYP2E1 CNV region is highlighted in (d) with a red box.

Multi-ethnic CYP450 copy number variation
S Martis et al

The Pharmacogenomics Journal (2013), 558–566 © 2013 Macmillan Publishers Limited
narrowed by breakpoint-spanning PCR and sequencing, which indicated that unequal crossing over most likely occurred at a 529-bp interval within introns 4 and exons 5 of CYP2B6 and CYP2B7P1 where the two genes are nearly identical. Additionally, both introns 4 contain four directly oriented homologous Alu elements (AluSc, AluSx, AluSx and AluSg/x) that may also facilitate NAHR at this locus.50 CYP2B6*30 is predicted to have 15 amino-acid changes, including the p.R487C alteration consistent with the *5 allele.1,51 More importantly, the exon 7 p.R378X non-sense alteration characteristic of the CYP2B6*28 allele is also encoded by *30, indicating that the translated *30 polypeptide would be truncated and lacking a heme binding site.52 Taken together, these amino-acid alterations indicate that the duplicated CYP2B6*30 allele is most likely non-functional. Detecting CNV alleles like CYP2B6*29 and *30 has significant pharmacogenetic implications when interpreting the metabolizer phenotype. For example, if just assessing copy number of exons 1 through 4, *30 could be misclassified as CYP2B6*1 × 2, which predicts an ultra-rapid metabolizer phenotype. Although the CYP2B6*29 (deletion) and *30 (duplication) alleles were rare in the tested populations (~1% in African-Americans and Asians, respectively), these variant alleles have been identified in other population studies as evidenced by entries in the DGV for this region.

CYP2E1 metabolizes clinically important drugs such as acetaminophen, isoniazid and tamoxifen, and is involved in the activation of several carcinogens. Additionally, CYP2E1 is the major component of the microsomal ethanol oxidizing system and accounts for ~10% of ethanol oxidation in the liver, indicating that this locus likely has an important role in alcoholism and related disorders. For example, some studies previously have associated variant CYP2E1 alleles (for example, *1D, *5B and *6) with alcoholism and cancer development in different ethno-racial groups.53–58 Notably, although the DGV indicates that the 10q26.3 region is copy number variable, no CYP2E1 CNV alleles previously have been characterized by the Cytochrome P450 Nomenclature Committee (http://www.cypalleles.ki.se/cyp2e1.htm).

Although we detected the CYP2E1 duplication (designated as CYP2E1*1Cx2) by the Cytochrome P450 Nomenclature Committee in all populations except Asians, no deletion alleles were identified in any of the studied racial or ethnic groups. High-resolution aCGH mapped the size of the CYP2E1 duplication to ~130 kb and localized the CNV breakpoints to two highly homologous directly oriented LCRs that share >98% sequence homology and include SPRN and its distal SPRN1 pseudogene. Of note, SPRN is a member of the prion protein family and variant SPRN alleles have recently been associated with Creutzfeldt-Jakob disease susceptibility.59 Additionally, our aCGH data indicate that the CYP2E1 duplication region also includes the SYCE1 gene. SYCE1 is recruited to the synaptosomal complex during meiosis and, interestingly, microdeletions that encompass SYCE1 recently have been associated with premature ovarian failure.60 Together, these data suggest that 10q26.3 CNVs may have additional clinical relevance beyond CYP2E1 pharmacogenetics and that adaptive selective pressures may play a role in why deletions in this region were not detected in any of our tested populations.

In conclusion, MLPA and TaqMan qPCR assays were used to determine the frequencies of deletion and duplication alleles for 11 CYP450 genes in a multi-ethnic healthy cohort. In addition to CYP2D6, CNVs were detected in CYP2A6, CYP2B6 and CYP2E1, and their frequencies varied among the studied populations. Of note, the discordant deletion and duplication allele frequencies for certain CYP450 CNV alleles (for example, CYP2A6 duplication and CYP2E1 deletion) suggest that selective pressures have likely acted on these alleles. The identified CYP2A6, CYP2B6 and CYP2E1 CNV alleles and their frequencies are strongly supported by the entries in the DGV (http://projects.tcag.ca/variation/) from other population studies. Importantly, our study also identified the novel CYP2B6/C287P1 duplicated fusion allele (designated as CYP2B6*30) and the CYP2E1 duplication (designated as CYP2E1*1Cx2). Moreover, high-resolution aCGH and sequencing analyses localized breakpoint regions to directly oriented LCRs with >90–98% homology, which indicate that these CYP450 CNV formations are likely mediated by NAHR resulting in both full gene and gene-fusion copy number imbalances. Detection of these CNVs should be considered when interrogating these genes for pharmacogenetic drug selection and dosing.

CONFLICT OF INTEREST

One of the authors (RV) is an employee of MRC-Holland, Amsterdam, The Netherlands. The remaining authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This research was supported in part by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH), through Grant KL2TR000069 (AS). The Cytochrome P-450 MLPA kit reagents used in this study were generously provided by MRC-Holland (Amsterdam, The Netherlands). We thank Dr Sarah Sim, Karolinska Institutet, Stockholm, Sweden, for critical reading of the manuscript; Ms Edith Gould, formerly of the Mount Sinai School of Medicine, New York, for her technical assistance with the MLPA assays; and Dr Minjie Luo, Mount Sinai School of Medicine, for assistance with the long-range PCR assays.

REFERENCES

1. Rovelet-Lecreux A, Hannequin D, Raux G, Le Meur N, Laquerriere A, Vital A et al. App locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet 2006; 38: 24–26.
2. McCarron SA, Huetta E, Kuballa P, Chlewski SD, Landry A, Goyette P et al. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn’s disease. Nat Genet 2008; 40: 1107–1112.
3. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature 2010; 466: 368–372.
4. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y et al. Detection of large-scale variation in the human genome. Nat Genet 2004; 36: 949–951.
5. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P et al. Large-scale copy number polymorphism in the human genome. Science 2004; 305: 525–528.
6. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU et al. Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 2005; 77: 78–88.
7. Pinto D, Marshall C, Feuk L, Scherer SW. Copy-number variation in control population cohorts. Hum Mol Genet 2007; 16 Spec No. 2: R168–R173.
8. Conrad DF, Pinto D, Redon R, Feuk L, Gokcen U, Zhang Y et al. Origins and functional impact of copy number variation in the human genome. Nature 2010; 464: 704–712.
9. Johansson AC, Feuk L. Characterization of copy number-stable regions in the human genome. Hum Mutat 2011; 32: 947–955.
10. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet 2006; 7: 85–97.
11. Hastings PJ, Lupski JR, Rosenberg SM, Isa G. Mechanisms of change in gene copy number. Nat Rev Genet 2009; 10: 551–564.
12. Zhang K, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. The DNA replication FoSTeS/MMBR mechanism can generate genomic, genic and exonic complex rearrangements in humans. Nat Genet 2009; 41: 849–853.
13. Hebbelinck SJ, Adjeti AA, Baer JL, Jenkins GD, Zhang J, Cunningham JM et al. Human SULT1A1 gene: copy number differences and functional implications. Hum Mol Genet 2007; 16: 463–470.
14. Huang RS, Chen P, Wisel S, Duan S, Zhan G, Cook EH et al. Population-specific GSTM1 copy number variation. Hum Mol Genet 2009; 18: 366–372.
15. Gaedigk A, Twist GP, Leeder JS. CYP2D6, SULT1A1 and UGT2B17 copy number variation. Functional and clinical implications. Pharmacogenomics 2012; 13: 91–111.
16. McGraw J, Waller D. Cytochrome CYP450 variations in different ethnic populations. Expert Opin Drug Metab Toxicol 2012; 8: 371–382.
17. He Y, Hoskins JM, McLeod HL. Copy number variants in pharmacogenetic genes. Trends Mol Med 2011; 17: 244–251.
18. Ramamoorthy A, Skaar TC. Gene copy number variations: it is important to determine which allele is affected. Pharmacogenomics 2011; 12: 299–301.

© 2013 Macmillan Publishers Limited
The Pharmacogenomics Journal (2013), 558 – 566
Nakajima M, Yoshida R, Fukami T, McLeod HL, Yokoi T. Novel human CYP2A6

Johansson I, Lundqvist E, Bertilsson L, Dahl ML, Sjoqvist F, Ingelman-Sundberg M.

Scott SA, Edelmann L, Kornreich R, Erazo M, Desnick RJ. CYP2C9, CYP2C19 and

Scott SA, Cohen N, Brandt T, Toruner G, Desnick RJ, Edelmann L. Detection of low-

Gaedigk A, Jaime LK, Bertino Jr JS, Berard A, Pratt VM, Bradfordand LD et al.

identification of novel CYP2D7-2D6 hybrids: non-functional and functional variants. Front Pharmacol 2010; 2: 121.

Fukami T, Nakajima M, Yamanaka H, Fukushima Y, McLeod HL, Yokoi T. A novel duplication type of CYP2A6 gene in African-American population. Drug Metab Dispos 2007; 35: 515–520.

Nakajima M, Yoshida R, Fukami T, McLeod HL, Yokoi T. Novel human CYP2A6 alleles confound gene deletion analysis. FEBS Lett 2004; 569: 75–81.

Johansson I, Lundqvist E, Bertilsson L, Dahl ML, Sjöqvist F, Ingelman-Sundberg M. Inherited amplification of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine. Proc Natl Acad Sci USA 1993; 90: 11825–11829.

Sachse C, Brockmoller J, Hildebrand M, Muller K, Roots I. Correctness of prediction of the CYP2D6 phenotype confirmed by genotyping 47 intermediate and poor metabolizers of debrisoquine. Pharmacogenet 1998; 8: 181–185.

Ishiguro A, Kubota T, Ishikawa H, Iga T. Metabolic activity of dextromethorphan O-demethylation in healthy Japanese volunteers carrying duplicated CYP2D6 genes: duplicated allele of CYP2D6*10 does not increase CYP2D6 metabolic activity. Clin Chim Acta 2004; 346: 201–204.

Scott SA, Jaremko M, Lubitz SA, Kornreich R, Halperin JL, Desnick RJ. CYP2C9*8 is prevalent among African-Americans: implications for pharmacogenetic dosing. Pharmacogenomics 2009; 10: 1243–1255.

Scott SA, Kashawneh R, Peter I, Kornreich R, Desnick RJ. Combined CYP2C9, VKORC1 and CYP4F2 frequencies among racial and ethnic groups. Pharmacogenomics 2010; 11: 781–791.

Martis S, Peter I, Hulot JS, Kornreich R, Desnick RJ, Scott SA. Multi-ethnic distribution of clinically relevant CYP2C2 genotypes and haplotypes. Pharmacogenomics J; e-pub ahead of print, 10 April 2012.

Scott SA, Edelmann L, Kornreich R, Erazo M, Desnick RJ. CYP2C9, CYP2C19 and CYP2D6 allele frequencies in the Ashkenazi Jewish population. Pharmacogenomics 2007; 8: 721–730.

Scott SA, Edelmann L, Kornreich R, Desnick RJ. Warfarin pharmacogenetics: CYP2C9 and VKORC1 genotypes predict different sensitivity and resistance frequencies in the Ashkenazi and Sephardi Jewish populations. Am J Hum Genet 2000; 67: 495–500.

Scott SA, Edelmann L, Kornreich R, Desnick RJ, Ciganda H, Desnick RJ. Identification of CYP2C19*2 and CYP2C19*3 genotypes in African-American and White populations. Am J Hum Genet 2009; 85: 847–853.

Scott SA, Ciganda H, Desnick RJ, Ciganda H. Genetic variability of CYP2C19 in populations of African and Asian origin: allele frequencies, novel functional variants, and possible implications for anti-HIV therapy with efavirenz. Pharmacogenet Genomics 2005; 15: 861–873.

Gatakan A, Hayashida T, Tsuchiya K, Yoshino M, Kuwahara T, Tsukada H et al. Successful efavirenz dose reduction in HIV type 1-infected individuals with cytochrome P450 2D6*6 and *26. Clin Infect Dis 2007; 45: 1230–1237.

Matzer MA, Deininger PL. Alu repeats and human genomic diversity. Nat Rev Genet 2002; 3: 370–379.

Hesse LM, He P, Krishnaswamy S, Hao Q, Hogan K, von Moltke LL et al. Pharmacogenetic determinants of interindividual variability in bupropion hydroxylation by cytochrome P450 2D6 in human liver microsomes. Pharmacogenetics; 2004; 14: 225–238.

Kircheiner J, Klein C, Meineke I, Sasse J, Zanger UM, Murdter TE et al. Bupropion and 4-OH-bupropion pharmacokinetics in relation to genetic polymorphisms in CYP2B6. Pharmacogenet Genomics 2003; 13: 619–626.

Rotger M, Tegude H, Colombo S, Cavassini M, Furrer H, Decosterd L et al. Predictive value of known and novel alleles of CYP2B6 for efavirenz plasma concentrations in HIV-infected individuals. Clin Pharmacol Ther 2007; 81: 557–566.

Grove J, Brown AS, Daly AK, Bassendine MF, James OF, Day CP. The Riai polymorphism of CYP2E1 and susceptibility to alcoholic liver disease in Caucasians: effect on age of presentation and dependence on alcohol dehydrogenase genotype. Pharmacogenet 1998; 8: 335–342.

Iwahashi K, Ameno S, Ameno K, Okada N, Kinoshita H, Sakae Y et al. Relationship between alcoholism and CYP2E1 C/D polymorphism. Neuropsychobiology 1998; 38: 218–221.

McCarver DG, Byan R, Hines RN, Hichme M, Wegene W. A genetic polymorphism in the regulatory sequences of human CYP2E1: association with increased chloroxazone hydroxylation in the presence of obesity and ethanol intake. Toxicol Appl Pharmacol 1998; 152: 276–281.

Howard LA, Ahluwalia JS, Lin SK, Sellers EM, Tandef RD. CYP2E1*1D regulatory polymorphism: association with alcohol and nicotine dependence. Pharmacogenetics 2003; 13: 321–328.

Khan AJ, Buwali M, Choudhuri G, Mathur N, Husain Q, Parmar D. Polymorphism in cytochrome P450 2E1 and interaction with other genetic risk factors and susceptibility to alcoholic liver cirrhosis. Mutat Res 2009; 664: 55–63.

Beck JA, Campbell TA, Adamson G, Poulter M, Uphill JB, Molou E et al. Association of a null allele of SPINK5 with variant Cretzfeldt-Jakob disease. J Med Genet 2008; 45: 813–817.

McGuire MM, Bowden W, Engel NJ, Ahn HW, Kovanci E, Rajkovic A. Genomic analysis using high-resolution single-nucleotide polymorphism arrays reveals novel microdeletions associated with premature ovarian failure. Fertil Steril 2011; 95: 1595–1600.

P450 2E1 genes: identification of a polymorphic repeat in the human gene. Biochim Biophys Res Comm 1999; 263: 286–293.

Tyndale RF, Sellers EM. Variable CYP2A6-mediated nicotine metabolism alters smoking behavior and risk. Drug Metab Dispos 2001; 29(4 Pt 2): 548–552.

Kamatana T, Fujieda M, Kiyotani K, Iwano S, Kunitoh H. Genetic polymorphism of CYP2A6 as one of the potential determinants of tobacco-related cancer risk. Biochim Biophys Res Comm 2005; 338: 306–310.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD et al. Global variation in copy number in the human genome. Nature 2006; 444: 444–454.

Rodriguez-Antona C, Gomez A, Karlgen M, Sim SC, Ingelman-Sundberg M. Molecular genetics and epigenetics of the cytochrome P450 2D gene family and its relevance for cancer risk and treatment. Hum Genet 2010; 127: 1–17.

Oscarson M, McLellan RA, Gullsten H, Yue YQ, Lang MA, Bernal ML et al. Characterisation and PCR-based detection of a CYP2A6 gene deletion found at a high frequency in a Chinese population. FEBS Lett 1999; 448: 105–110.

van Binsbergen E. Origins and breakpoint analyses of copy number variations: up close and personal. Cytogenet Genome Res 2011; 135: 271–276.

Mwenifumbo JC, Zhou Q, Benowitz NL, Sellers EM, Tyndale RF. New CYP2A6 gene deletion and conversion variants in a population of Black African descent. Pharmacogenomics 2010; 11: 189–198.

Lang T, Klein C, Fischer J, Nussler AK, Neuhaus P, Hoffmann U et al. Extensive genetic polymorphism in the human CYP2B6 gene with impact on expression and function in human liver. Pharmacogenomics 2001; 11: 399–415.

Klein C, Lang T, Sausselle T, Barbosa-Scard E, Schunck WH, Eichelbaum M et al. Genetic variability of CYP2B6 in populations of African and Asian origin: allele frequencies, novel functional variants, and possible implications for anti-HIV therapy with efavirenz. Pharmacogenet Genomics 2005; 15: 861–873.

Gatagan A, Hayashida T, Tsuchiya K, Yoshino M, Kuwahara T, Tsukada H et al. Successful efavirenz dose reduction in HIV type 1-infected individuals with cytochrome P450 2D6*6 and *26. Clin Infect Dis 2007; 45: 1230–1237.

Patzer MA, Deininger PL. Alu repeats and human genomic diversity. Nat Rev Genet 2002; 3: 370–379.