Structural variation at the *CYP2C* locus: Characterization of deletion and duplication alleles

Mariana R. Botton[1,2,*], **Xingwu Lu**[1,2], **Geping Zhao**[2], **Elena Repnikova**[3,4], **Yoshinori Seki**[2], **Andrea Gaedigk**[4,5], **Eric E. Schadt**[1,2], **Lisa Edelmann**[1,2], **Stuart A. Scott**[1,2,†]

[1]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

[2]Sema4, a Mount Sinai venture, Stamford, CT 06902, USA.

[3]Clinical Genetics and Genomics Laboratories, Children's Mercy Hospital Kansas City, Kansas City, MO 64108, USA

[4]School of Medicine, University of Missouri-Kansas City, Kansas City, MO 64108, USA.

[5]Division of Clinical Pharmacology, Toxicology & Therapeutic Innovation, Children's Mercy Kansas City, Kansas City, MO 64108, USA.

Abstract

The human *CYP2C* locus harbors the polymorphic *CYP2C18*, *CYP2C19*, *CYP2C9* and *CYP2C8* genes, and of these, *CYP2C19* and *CYP2C9* are directly involved in the metabolism of ~15% of all medications. All variant *CYP2C19* and *CYP2C9* star (*) allele haplotypes currently catalogued

[†]**Correspondence to: Stuart A. Scott, PhD**, Associate Professor, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1497, New York, NY, 10029, Tel. 212-241-3780, Fax. 212-241-0139, stuart.scott@mssm.edu.

[*]Current address: Experimental Research Center, Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, Rio Grande do Sul, Brazil; Graduate Program in Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil.

by the Pharmacogene Variation (PharmVar) Consortium are defined by sequence variants. To determine if structural variation also occurs at the *CYP2C* locus, the 10q23.33 region was interrogated across deidentified clinical chromosomal microarray (CMA) data from 20,642 patients tested at two academic medical centers. Fourteen copy number variants that affected the coding region of *CYP2C* genes were detected in the clinical CMA cohorts, which ranged in size from 39.2–1,043.3 kb. Selected deletions and duplications were confirmed by MLPA or ddPCR. Analysis of the clinical CMA and an additional 78,839 cases from the Database of Genomic Variants (DGV) and ClinGen (total n=99,481) indicated that the carrier frequency of a *CYP2C* structural variant is ~1 in 1000, with ~1 in 2,000 being a *CYP2C19* full-gene or partial-gene deletion carrier, designated by PharmVar as *CYP2C19*36* and *37*, respectively. Although these structural variants are rare in the general population, their detection will likely improve metabolizer phenotype prediction when interrogated for research and/or clinical testing.

Keywords

CYP2C; *CYP2C19*; *CYP2C9*; copy number variation; deletion; duplication; pharmacogenomics; chromosomal microarray; database

INTRODUCTION

The human cytochrome P450 (CYP) enzyme superfamily is responsible for the oxidative metabolism of many drugs, xenobiotics, and other endogenous substances. The polymorphic *CYP2C* locus at chromosome 10q23.33 is comprised of the *CYP2C18*, *CYP2C19*, *CYP2C9* and *CYP2C8* genes, which encode enzymes that together are involved in the hepatic metabolism of ~25% of commonly prescribed drugs. Moreover, CYP2C19 and CYP2C9 metabolize ~15% of drugs currently listed on the U.S. Food and Drug Administration (FDA) Pharmacogenomic Biomarkers in Drug Labeling table (www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm, which includes medications commonly used in neurology, rheumatology, psychiatry, cardiology, gastroenterology, gynecology, and infectious disease.

Currently 35 *CYP2C19* and 60 *CYP2C9* variant star (*) allele haplotypes are catalogued in the Pharmacogene Variation Consortium (PharmVar) database (www.pharmvar.org) (Gaedigk et al., 2018; Gaedigk et al., 2019). Without exception, all are defined by sequence variants. As such, no *CYP2C19* or *CYP2C9* star (*) alleles currently include structural variation (e.g., copy number variants (CNV)), which is consistent with our previously reported pilot study that employed multiplex ligation-dependent probe amplification (MLPA) screening of this gene region across a multi-ethnic cohort of ~500 individuals (Martis et al., 2013). However, pharmacogenomic CNV alleles can play important roles in enzyme activity and drug response variability (He, Hoskins, & McLeod, 2011), which have been characterized among some cytochrome P450 (*CYP2B6*, *CYP2D6*), glutathione S-transferase (*GSTT1*, *GSTM1*), and sulfotransferase (*SULT1A1*, *SULT2A1*) genes (Gaedigk, Gaedigk, & Leeder, 2010; Gaedigk, Twist, & Leeder, 2012; Gjerde et al., 2008; Martis et al., 2013; Schulze et al., 2013; Vijzelaar et al., 2018). Notably, interrogation of publicly available sequencing data has recently indicated that some populations harbor CNV alleles

at the *CYP2C* gene region (Santos et al., 2018). The potential clinical significance of low frequency structural variation at the *CYP2C* gene region in the general population prompted our interrogation of chromosomal microarray (CMA) data from multiple database sources for pharmacogenomic CNV discovery.

DATA SPECIFICATIONS

Data type	Tables of copy number variant (CNV) coordinates
Data acquisition method	Chromosomal microarray (CMA)
Data format	Analyzed
Experimental factors	Germline DNA tested by CMA for clinical indication or involvement in population genomics research
Experimental features	*CYP2C* gene region [chr10:96100000_97100000 (GRCh37/hg19)] was interrogated across CMA data from 99,481 individuals
Data source location	New York, NY, Kansas City, MO, publicly available CNV databases
Data accessibility	Data in this study are available at public repositories: Leiden Open Variation Database (LOVD) (https://databases.lovd.nl; DOI: 10.1002/humu.21438) cases: https://databases.lovd.nl/shared/individuals/00239123 https://databases.lovd.nl/shared/individuals/00239124 https://databases.lovd.nl/shared/individuals/00239125 https://databases.lovd.nl/shared/individuals/00239126 https://databases.lovd.nl/shared/individuals/00239127 https://databases.lovd.nl/shared/individuals/00239128 https://databases.lovd.nl/shared/individuals/00239129 https://databases.lovd.nl/shared/individuals/00239130 https://databases.lovd.nl/shared/individuals/00239131 https://databases.lovd.nl/shared/individuals/00239132 https://databases.lovd.nl/shared/individuals/00239133 https://databases.lovd.nl/shared/individuals/00239134 https://databases.lovd.nl/shared/individuals/00239135 https://databases.lovd.nl/shared/individuals/00239136 https://databases.lovd.nl/shared/individuals/00239137 Clinical Genome Consortium (http://dbsearch.clinicalgenome.org/search/; DOI: 10.1056/NEJMsr1406261) Database of Genomic Variants (http://dgv.tcag.ca; DOI: 10.1093/nar/gkt958) DECIPHER (https://decipher.sanger.ac.uk; DOI: 10.1016/j.ajhg.2009.03.010) Pharmacogene Variation Consortium (PharmVar) database (www.pharmvar.org; DOI: 10.1002/cpt.1268)

IMPACT OF DATA

In contrast to the well-described pharmacogenomic alleles that are relatively common in the general population (e.g., *CYP2C9*2, CYP2C19*2*), it is increasingly appreciated that the majority of human genetic variation is actually rare [minor allele frequency (MAF) <1%] (Genomes Project et al., 2010; Tennessen et al., 2012), making association studies between these variants and drug response phenotypes challenging (Verma et al., 2018). To facilitate the discovery of low frequency variants that potentially influence drug response, recent studies have interrogated high-throughput sequencing data across drug target genes (Nelson et al., 2012), *CYP450* genes (Gordon et al., 2014), and selected drug absorption, distribution, metabolism and excretion (ADME) and other candidate pharmacogenes (Bush et al., 2016; Li et al., 2014; Santos et al., 2018), all indicating that rare and potentially functional pharmacogenomic variants are prevalent in diverse populations. These studies highlight the importance of interrogating large datasets for rare pharmacogenomic variation

discovery, which prompted our interrogation of deidentified CMA data from large databases to further study low frequency pharmacogenomic structural variation at the clinically relevant *CYP2C* gene cluster region.

MATERIALS AND METHODS

Experimental Design

The design of this study included interrogation of multiple databases of CMA data, including both clinical cytogenomic laboratory CMA data (discovery and replication cohorts), as well as publicly available CMA data from the general population (research and clinical cohorts). Selected *CYP2C* gene deletion and duplication samples from the discovery and replication cohorts were analytically confirmed by orthogonal copy number methods.

Clinical Cytogenomic Testing Cohort – Discovery

Individuals in the discovery cohort were referred to the Cytogenetics and Cytogenomics Laboratory at Mount Sinai Genomics Inc. (DBA Sema4), New York, from 2010 to 2017 for pre- or postnatal clinical CMA testing. All patients were tested with informed consent and deidentified CMA data were stored in an internal database that enabled interrogation and CNV frequency analyses. A total of 11,096 unique samples were analyzed, which included 6,083 prenatal (amniotic fluid, chorionic villus specimens, fetal blood) and 5,013 postnatal (peripheral blood, products of conception) samples. Although race and ethnicity were not commonly available, the prenatal cohort self-reported as white (82.5%), Asian (8.0%), black (4.7%), Hispanic/Latino (4.5%) and American Indian (0.36%), and the postnatal cohort self-reported as white (59.1%), Hispanic/Latino (19.1%), black (11.8%), Asian (6.3%), and American Indian (3.8%).

Clinical Cytogenomic Testing Cohort – Replication

Individuals in the replication cohort were referred to the Clinical Genetics and Genomics Laboratory at Children's Mercy, Kansas City, from 2009 to 2018 for pre- or postnatal clinical CMA testing. All patients were tested with informed consent and deidentified CMA data were stored in an internal database that enabled interrogation and CNV frequency analyses. A total of 9,760 unique peripheral blood samples were analyzed as a replication cohort for structural variation at the *CYP2C* locus. Race and ethnicity demographics of the patient cohort were not available.

Chromosomal Microarray (CMA) Analysis – Discovery

CMA was performed on the discovery cohort using the Agilent Technologies platform (Santa Clara, CA, USA) according to the manufacturer's instructions and as previously reported (Reiner et al., 2017; Scott et al., 2010). Throughout the period of clinical CMA testing and data analysis (2010–2017), three commercial microarrays were used that had increasing probe density and resolution [44K (design 015141), 105K (design 031750), and 180K (design 029830); Agilent Technologies]; however, all three microarray designs had adequate probe coverage across the *CYP2C* gene cluster region to detect multi-exon CNVs within *CYP2C18* (NG_008373.1), *CYP2C19* (NG_008384.3), *CYP2C9* (NG_008385.1),

and/or *CYP2C8* (NG_007972.1) (Figure 1). All genomic coordinates are reported using NCBI human genome reference Build 37 (GRCh37/hg19).

Chromosomal Microarray (CMA) Analysis – Replication

CMA was performed on the replication cohort using the Affymetrix Cytoscan® HD CNV +SNP array platform (Santa Clara, CA, USA) or the Agilent 244K whole genome oligonucleotide microarray (design 014693; Santa Clara, CA, USA) according to the manufacturer's instructions. The CytoScan® HD microarray contains 1,953,246 non-polymorphic and 743,304 single nucleotide polymorphism (SNP) markers, which are enriched in disease gene areas, and the Agilent 244K microarray contains ~244,000 oligonucleotide probes spaced at a median distance of 6.4 kb across the human genome (Figure 1). Data were analyzed by ChAS 3.2 (Affymetrix) or Genomic Workbench (Agilent Technologies) software as appropriate. As above, all genomic coordinates are reported using NCBI human genome reference Build 37 (GRCh37/hg19).

Copy Number Variation (CNV) Confirmation

Multiplex Ligation-dependent Probe Amplification (MLPA)—Copy number results from CMA testing were validated by multiplex ligation-dependent probe amplification (MLPA) testing on samples from the discovery cohort with available DNA. MLPA was performed using the Cytochrome P-450 MLPA kit (P128-B1; MRC-Holland, Amsterdam, The Netherlands) according to the manufacturer's instructions and as previously reported (Martis et al., 2013; Vijzelaar et al., 2018). This commercial MLPA probe mix includes three *CYP2C19* probes (exons 2, 6, and 9) and five *CYP2C9* probes (exons 2, 7, 8 [2 probes], and 9) (Figure 1), plus an additional 34 probes that interrogate 12 other pharmacogenetic genes (*CYP2D6, CYP1B1, CYP3A4, CYP3A5, CYP2E1, CYP1A1, CYP1A2, CYP2A6, CYP2B6, GSTP1, GSTT1* and *GSTM1*) (Martis et al., 2013). Amplified products were separated by capillary gel electrophoresis and analyzed using GeneMarker v1.90 software (SoftGenetics, State College, PA). After quality control and data normalization, copy number was determined according to the following peak ratio ranges: one copy >0.25 and <0.75; two copies >0.75 and <1.25; three copies >1.25 and <1.7; four copies >1.7 and <2.0.

Droplet Digital PCR (ddPCR)—Copy number results from CMA testing were also validated by droplet digital PCR (ddPCR) testing on samples from the replication cohort with available DNA. TaqMan™ copy number assays targeting *CYP2C19* exon 2 (Hs05148033_cn) and intron 6 (Hs02932336_cn) were employed and signals normalized against the TERT gene (Cat# 4403316; Thermo Fisher, Waltham, MA) (Figure 1), and analysis was performed using the Bio-Rad QX-200 Droplet Digital PCR System (Bio-Rad, Hercules, CA). Genomic DNA were digested with EcoRI-HF (New England BioLabs, Ipswich, MA) and inactivated at 65ºC. Digested DNA were subsequently combined with 1X ddPCR Supermix for Probes (Bio–Rad, Hercules, CA), TaqMan™ and TERT reference assays. Droplets were generated with the Auto Droplet Generator and cycled in a C1000 Touch Thermocycler using recommended parameters. Droplets were analyzed with the QX200 Droplet Reader instrument and data analysis performed with the Quantasoft™ Software (Bio-Rad, Hercules, CA).

Cytogenomic Copy Number Variation (CNV) Population Cohorts

Structural variation databases were also interrogated to identify *CYP2C* region CNVs in healthy and clinical cohorts. Three independent sources were utilized: (1) the Database of Genomic Variants (DGV) (http://dgv.tcag.ca), which catalogues structural variation (>50 bp) in healthy population samples; (2) Clinical Genome Consortium (ClinGen) (http://dbsearch.clinicalgenome.org/search/), which catalogues genomic variation in patient population samples; and (3) DECIPHER (https://decipher.sanger.ac.uk), which also catalogues genomic variation in patient population samples. A 1 Mb genomic region was queried across all public databases [chr10:96100000_97100000 (GRCh37/hg19)], and only CNVs that included coding regions of any *CYP2C* gene were included in the study. Larger overlapping chromosome 10q23.33-q24.1 deletions and duplications (>2 Mb) in the clinical databases were excluded from the *CYP2C* CNV allele and carrier frequency analyses, as these likely pathogenic aberrations would be more consistent with a syndromic Mendelian phenotype.

DATA:

Chromosomal Microarray (CMA) Detection of *CYP2C* Deletions and Duplications

The CMA probe coverage across the *CYP2C* region for all clinical microarrays used in the study are illustrated in Figure 1. In the discovery cohort (ISMMS/Sema4; n=11,096), *CYP2C* gene region CNVs were detected in nine unrelated patients, including seven deletions and two duplications (Table 1 and Table 2). The identified deletions ranged in size from 52.0 kb (exons 8 to 9 of *CYP2C18* and exons 1 to 4 of *CYP2C19*) to 421.0 kb (including *TBC1D12, HELLS, CYP2C18,* and *CYP2C19*) (Table 1 and Figure 1). The identified duplications were larger, 663.8 kb and 1.0 Mb, and included all *CYP2C* (*CYP2C18*, *CYP2C19*, *CYP2C9*, *CYP2C8*) and the neighboring *ACMS6*, *PDLIM1* and *SORBS1* genes (Table 1 and Figure 1). In the replication cohort (CMH; n=9,760), *CYP2C* region CNVs were detected in six unrelated patients, including three *CYP2C* deletions and three *CYP2C* duplications (Table 1 and Table 2). The deletions ranged in size from 39.2 kb to 61.5 kb (exons 1 to 5 of *CYP2C19*) (Table 1 and Figure 1), whereas the duplications (130.8 and 131.9 kb) included the 3' region of *CYP2C19* (exon 9 with or without exon 8) and exons 1 to 7 of *CYP2C9* (Table 1 and Figure 1). All *CYP2C* CNV alleles identified in the discovery and replication CMA cohorts were submitted to the Leiden Open Variation Database (LOVD) (https://databases.lovd.nl) (Fokkema et al., 2011), and their unique variant IDs are listed in Table 1.

Confirmation of *CYP2C* Copy Number Variants (CNVs)

Among all subjects with *CYP2C* CNVs identified by CMA testing, five samples from the discovery cohort had available DNA for confirmation by MLPA testing. The locations of the MLPA probes in relation to the CMA probes are illustrated in Figure 1. All MLPA results were consistent with the CNVs detected by CMA testing (Table 2). Of note, given that the CMA and MLPA platforms have unique probe locations to interrogate copy number across the *CYP2C* region, deletions that affected only *CYP2C18* and/or only exon 1 of *CYP2C19* were not detected by MLPA. As noted in the *Materials and Methods*, the MLPA probe mix only interrogates exons 2, 6, and 9 of *CYP2C19* and exons 2, 7, 8, and 9 of *CYP2C9* (Figure

1). In addition, all subjects with *CYP2C19* CNVs identified by CMA testing in the replication cohort were confirmed by ddPCR. The locations of the ddPCR probes in relation to the CMA probes are illustrated in Figure 1. All ddPCR results at exon 2 and intron 6 of *CYP2C19* were consistent with the partial gene deletions and duplications detected by CMA testing.

Database Detection of *CYP2C* Copy Number Variants (CNVs)

Structural variants within a 1 Mb region at the *CYP2C* locus (chr10:96100000_97100000) were also identified in the DGV, which is commonly considered to be a CMA database representative of the general population. These *CYP2C* CNVs were consistent with those detected in our clinical cohorts and are illustrated in Figure 1 and detailed in the Appendix. A total of 36 CNVs (from 9 independent studies) overlapping the *CYP2C18*, *CYP2C19*, *CYP2C9*, and/or *CYP2C8* genes were catalogued in the DGV. The majority (n=27; 75%) were deletions, including 17 and 4 deletions that overlapped coding regions of *CYP2C19* and/or *CYP2C9*, respectively (Table 3).

Structural variants at the *CYP2C* gene region were also identified in the ClinGen and DECIPHER databases. Given that these databases are comprised of CMA data from patient cohorts with variable phenotypes, larger deletions and duplications of the chromosome 10q23.33-q24.1 region were present; however, aberrations >2 Mb were not included in the *CYP2C* CNV analyses as detailed in the *Materials and Methods* (Appendix). Consistent with the DGV *CYP2C* CNVs, 22 CNVs were present in ClinGen and DECIPHER that overlapped *CYP2C18*, *CYP2C19*, *CYP2C9* and/or *CYP2C8*, including 13 deletions (8.6 kb to 969.7 kb), and eight duplications (8.9 kb to 1.1 Mb) (Figure 1, Table 3 and Appendix). Notably, the most common *CYP2C* CNVs in all three population databases were deletions that overlapped *CYP2C19* (n=26; 44.8% of all *CYP2C* CNVs).

CYP2C Deletion and Duplication Frequencies and Allele Nomenclature

The *CYP2C* CNV frequency data from our clinical CMA cohorts and the CNV population databases are summarized in Table 3. The *CYP2C* CNV data from DECIPHER was not incorporated into the population frequencies given the difficulty with determining an accurate size of this dynamic clinical cohort. Taken together, the overall carrier frequency of a *CYP2C* CNV is ~1 in 1000 [0.085% (95% CI: 0.067–0.104%)] (Table 3). The data for these CNV alleles were reviewed by the PharmVar Consortium (Gaedigk et al., 2018; Gaedigk et al., 2019), which subsequently classified a full gene *CYP2C19* deletion as *CYP2C19*36* and a partial gene *CYP2C19* deletion (that includes at least exon 1) as *CYP2C19*37* [combined carrier frequency of ~1 in 2000; 0.046% (95% CI: 0.033–0.059%)]. The full gene and partial gene *CYP2C9* deletion alleles identified in the DGV and ClinGen had a carrier frequency of ~1 in 20,000; 0.005% (95% CI: 0.001–0.009%)] (Table 3 and Figure 1). The *CYP2C19*, *CYP2C9*, and *CYP2C8* duplications did not receive designated star (*) alleles as PharmVar recommends classifying gene duplications based on their haplotype sequence and total detected copy number (e.g., *CYP2C9*1/*1x2*), consistent with established *CYP2D6* duplication nomenclature (Gaedigk et al., 2018) (www.pharmvar.org). Notably, one sample in the clinical CMA cohort with the largest chromosome 10q23.33 duplication (1043.3 kb) that included multiple *CYP2C* genes was

genotyped for a panel of *CYP2C* star (*) allele variants, which resulted in the following diplotypes: *CYP2C19*17/*17x2*, *CYP2C9*1/*1x2*, and *CYP2C8*1/*1x2*.

DISCUSSION

Recent studies that utilized large high-throughput sequencing datasets for pharmacogenomic variant discovery prompted our interrogation of deidentified CMA data from 99,695 individuals for pharmacogenomic CNV discovery at the clinically relevant *CYP2C* gene cluster region. Consistent with the sequencing studies that identified novel coding variants in pharmacogenomic genes (Bush et al., 2016; Gordon et al., 2014; Li et al., 2014; Nelson et al., 2012), our large CNV study resulted in the discovery of pharmacogenomic deletion and duplication alleles at the *CYP2C* region. The full gene and partial gene *CYP2C19* deletions have been designated by the PharmVar Consortium as *CYP2C19*36* and *CYP2C19*37*, respectively (www.pharmvar.org). Of note, the *CYP2C19*37* partial deletion allele is consistent with the *CYP2C19* partial gene deletions recently identified in the Exome Aggregation Consortium (ExAC) database (Ruderfer et al., 2016; Santos et al., 2018). Dissemination of these newly defined *CYP2C19* alleles through the widely utilized PharmVar database will likely increase awareness of these low frequency structural variants across the pharmacogenomics community.

The significance of structural variation in human disease and phenotypic diversity is increasingly being recognized, and several genomic studies have generated catalogs of CNVs to facilitate a better understanding of their clinical relevance (Johansson & Feuk, 2011; Sudmant et al., 2015). It is estimated that up to 60% of the human genome may contain structural variants in the general population, which typically range in size from 100 bp to 50 kb (Escaramis, Docampo, & Rabionet, 2015), and clinical interpretation of these aberrations when identified by CMA testing is facilitated by professional medical genetics practice guidelines (South et al., 2013). Larger gene-dense aberrations are more likely to result in penetrant syndromic phenotypes; however, some smaller CNVs have increasingly been implicated as susceptibility alleles for several phenotypes, including neurodegenerative disorders, cancer, autism, and psychiatric diseases (Cook & Scherer, 2008; Gonzalez et al., 2005; Han et al., 2017; Nishioka et al., 2006; Rovelet-Lecrux et al., 2006; Sebat et al., 2007). CNVs can influence these human traits by altering the copy number of dosage-sensitive genes (Douglas et al., 2005; Roa, Garcia, & Lupski, 1991) and/or modulating local gene expression (Cahan, Li, Izumi, & Graubert, 2009; Henrichsen, Chaignat, & Reymond, 2009).

Pharmacogenomic structural variation has been previously characterized at several clinically relevant regions (Santos et al., 2018), including *CYP450* genes (*CYP2A6*, *CYP2B6*, *CYP2C* cluster, *CYP2D6*), glutathione S-transferases (*GSTT1*, *GSTM1*), and sulfotransferases (*SULT1A1*, *SULT2A1*) (Gaedigk et al., 2010; Gjerde et al., 2008; Martis et al., 2013; Schulze et al., 2013; Vijzelaar et al., 2018). Notably, individuals with greater than two functional *CYP2D6* copies (e.g. **1xN*, **2xN*, **35xN*) have higher enzyme activity, whereas *CYP2D6*5* deletion alleles do not encode a functional CYP2D6 protein. Moreover, common full gene *GSTT1* and *GSTM1* deletions have been associated with increased risk for

chemotherapy toxicity among lymphoma patients (Cho et al., 2010) and susceptibility to tacrine hepatotoxicity among Alzheimer's patients (Simon et al., 2000).

Loss-of-function *CYP2C19* and *CYP2C9* alleles have extensive evidence for important roles in clopidogrel, voriconazole, antidepressant, warfarin and/or phenytoin response, which prompted recent Clinical Pharmacogenetics Implementation Consortium (CPIC) and Dutch Pharmacogenetics Working Group (DPWG) practice guidelines for *CYP2C19* and *CYP2C9* pharmacogenetic-guided prescribing (Caudle et al., 2014; Hicks et al., 2015; Hicks et al., 2017; Johnson et al., 2017; Moriyama et al., 2017; Scott et al., 2013; Swen et al., 2011). Given that the newly defined *CYP2C19*36* and **37* deletion alleles are presumed to be nonfunctional, it is expected that they would have the same effects on drug response as the well-known alleles defined by sequence variants (e.g., *CYP2C19*2, *3*). Interestingly, *CYP2C19* deletions have previously been reported in a Northern Finnish case-control study, which identified an association between *CYP2C19* deletion and triple-negative breast cancer (ER/PR/HER2 negative), implicating CYP2C19 in estrogen catabolism (Tervasmaki, Winqvist, Jukkola-Vuorinen, & Pylkas, 2014). Our combined analysis of clinical CMA data from two academic medical centers and publicly available databases indicate that the carrier frequency of a *CYP2C* structural variant is ~1 in 1000, with ~1 in 2000 being a *CYP2C19* deletion carrier; however, these aberrations may have higher allele frequencies in specific subpopulations (Santos et al., 2018; Tervasmaki et al., 2014).

Recurrent CNVs are often flanked by segmental duplications, which can act as substrates for non-allelic homologous recombination (NAHR) and the meiotic formation of both deletion and duplication alleles (Carvalho & Lupski, 2016). Importantly, two pairs of segmental duplications are nested within the *CYP2C* gene cluster region, one directly oriented ~10–20 kb element (~92% identical) that flank *CYP2C18* and *CYP2C19*, and a smaller element (~1.6 kb) directly oriented on the negative strand that are located at the 3' region of *CYP2C8* (Figure 1). In addition to these segmental duplications, the *CYP2C* subfamily genes have a high degree of sequence homology. This homology and the segmental duplications flanking *CYP2C18* and *CYP2C19* are likely driving the formation of the more common and recurrent deletions at this region, which is consistent with the NAHR mechanism generally favoring deletions over duplications (Liu, Carvalho, Hastings, & Lupski, 2012).

Notably, the resolution of microarrays used for CMA testing was different across platforms and laboratories. As such, our reported CNV sizes are based on the minimum number of probes that detected gains or losses by a specific microarray; however, based on both our data and those from public databases there are recurrent deletions that affect both *CYP2C18* and *CYP2C19*, as well as larger CNVs that can include multiple *CYP2C* genes. Unfortunately, precise breakpoints were not feasible to determine given the multiple CMA platforms used across studies, as well as the paucity of available DNA for follow up sequencing.

In conclusion, our interrogation of CMA data from almost 100,000 individuals identified low frequency pharmacogenomic CNVs at the clinically relevant *CYP2C* region in the general population. These results are consistent with previously reported pharmacogenomic sequencing studies, which identified a spectrum of rare pharmacogenomic variants that are

likely to be functional (Bush et al., 2016; Gordon et al., 2014; Li et al., 2014; Nelson et al., 2012). Although the identified *CYP2C* deletion and duplication alleles have low frequencies in the studied populations, their contribution to an individual's CYP2C19 and CYP2C9 metabolizer phenotype status is most likely clinically relevant. The nonfunctional deletion alleles can lead to either intermediate or poor metabolizer phenotypes, and the larger *CYP2C* gene region duplications discovered in our study could lead to an ultrarapid metabolizer phenotype across CYP2C19, CYP2C19, CYPC9 and CYP2C8 if the duplication alleles do not harbor sequence variants that obliterate function. These rare individuals would likely be at risk for atypical CYP2C-mediated metabolism across multiple drugs and drug classes (e.g., voriconazole, clopidogrel, phenytoin, fosphenytoin, phenobarbital, amytriptiline, torsemide). Although the technical infrastructure and additional cost of interrogating copy number at the *CYP2C* gene region may not currently be feasible for clinical laboratories that offer pharmacogenetic testing, the clinical relevance of these low frequency CNV alleles indicates that future iterations of clinical pharmacogenomic sequencing assays that incorporate computational copy number detection pipelines should include this gene family in addition to the pharmacogenes with more common CNV alleles.

ACKNOWLEDGEMENTS

APPENDIX

Appendix Table 1.

CYP2C copy number variants (CNVs) identified in the Database of Genomic Variants (DGV), ClinGen, and DECIPHER.

CNV ID	Genes	CNV type	Region size	Frequency	Study population	Ref.
DGV						
esv3624253	*CYP2C18*	Deletion	118,150	1/2504	Healthy	(Genomes Project et al., 2015)
esv2672446	*CYP2C18*	Deletion	118,136	1/1092	Healthy	(Genomes Project et al., 2012)
esv3624254	*CYP2C18*	Deletion	79,667	3/2504	Healthy	(Genomes Project et al., 2015)
dgv955n100	*CYP2C18*	Deletion	111,509	6/29084	Intellectual disability and/or developmental delay	(Coe et al., 2014)
nsv551954	*CYP2C18*	Deletion	115,921	2/15767	Intellectual disability and/or developmental delay	(Cooper et al., 2011)
dgv956n100	*CYP2C18*	Deletion	69,696	2/29084	Intellectual disability and/or developmental delay	(Coe et al., 2014)
esv2657299	*CYP2C18*	Deletion	8,618	1/1092	Healthy	(Genomes Project et al., 2012)
nsv522538	*CYP2C18, CYP2C19*	Deletion	142,135	1/2026	Healthy	(Shaikh et al., 2009)
dgv1355n54	*CYP2C18, CYP2C19*	Deletion	246,123	3/15767	Intellectual disability and/or developmental delay	(Cooper et al., 2011)
esv2674280	*CYP2C18, CYP2C19*	Deletion	155,809	1/1092	Healthy	(Genomes Project et al., 2012)
nsv1046308	*CYP2C18, CYP2C19*	Deletion	191,713	1/29084	Intellectual disability and/or developmental delay	(Coe et al., 2014)
dgv1355n54	*CYP2C18, CYP2C19*	Deletion	246,123	3/15767	Intellectual disability and/or developmental delay	(Cooper et al., 2011)
dgv957n100	*CYP2C18, CYP2C19*	Deletion	70,440	5/29084	Intellectual disability and/or developmental delay	(Coe et al., 2014)
esv3624256	*CYP2C18, CYP2C19*	Deletion	44,223	2/2504	Healthy	(Genomes Project et al., 2015)
esv2761617	*CYP2C18, CYP2C19*	Deletion	53,597	1/1109	Healthy	(Vogler et al., 2010)
esv3624258	*CYP2C18, CYP2C19*	Duplication	80,151	1/2504	Healthy	(Genomes Project et al., 2015)
dgv160e214	*CYP2C18, CYP2C19*	Deletion	80,151	1/2504	Healthy	(Genomes Project et al., 2015)
dgv1356n54	*CYP2C18, CYP2C19*	Deletion	61,544	6/15767	Intellectual disability and/or developmental delay	(Cooper et al., 2011)
esv3891886	*CYP2C18, CYP2C19, CYP2C9*	Deletion	271,380	1/3017	Infectious diseases, Thyrotoxic Hypokalemic Periodic Paralysis (THPP) and Hb E/b-thalassemia	(Suktitipat et al., 2014)
nsv1035409	*CYP2C18, CYP2C19, CYP2C9, CYP2C8*	Deletion	336,881	1/29084	Intellectual disability and/or developmental delay	(Coe et al., 2014)
nsv551962	*CYP2C19*	Deletion	54,075	1/15767	Intellectual disability and/or developmental delay	(Cooper et al., 2011)
esv2659638	*CYP2C19*	Deletion	61,848	1/1092	Healthy	(Genomes Project et al., 2012)
esv3624260	*CYP2C19*	Deletion	39,678	2/2504	Healthy	(Genomes Project et al., 2015)
nsv516555	*CYP2C19*	Deletion	12,880	2/2026	Healthy	(Shaikh et al., 2009)
nsv523259	*CYP2C19*	Duplication	159,144	1/2026	Healthy	(Shaikh et al., 2009)
nsv1052578	*CYP2C19*	Deletion	103,868	1/29084	Intellectual disability and/or developmental delay	(Coe et al., 2014)
nsv1047782	*CYP2C19, CYP2C9*	Duplication	137,386	1/29084	Intellectual disability and/or developmental delay	(Coe et al., 2014)
esv2741937	*CYP2C19, CYP2C9*	Deletion	138,826	1/96	Healthy	(Wong et al., 2013)
esv3624265	*CYP2C9*	Duplication	72,590	1/2504	Healthy	(Genomes Project et al., 2015)

CNV ID	Genes	CNV type	Region size	Frequency	Study population	Ref.
nsv551963	CYP2C9	Deletion	23,682	1/15767	Intellectual disability and/or developmental delay	(Cooper et al., 2011)
nsv7497	CYP2C9	Duplication	34,308	1/8	Healthy	(Kidd et al., 2008)
esv3624264	CYP2C9, CYP2C8	Duplication	175,419	1/2504	Healthy	(Genomes Project et al., 2015)
nsv1050544	CYP2C8	Duplication	47,756	1/29084	Intellectual disability and/or developmental delay	(Coe et al., 2014)
nsv1044182	CYP2C8	Duplication	30,694	1/29084	Intellectual disability and/or developmental delay	(Coe et al., 2014)
esv2761618	CYP2C8	Duplication	30,682	1/1109	Healthy	(Vogler et al., 2010)
nsv551964	CYP2C8	Deletion	56,916	1/15767	Intellectual disability and/or developmental delay	(Cooper et al., 2011)
nsv467436	CYP2C8	Deletion	56,916	1/2493	Healthy	(Itsara et al., 2009)
ClinGen						
Pathogenic						
nssv13651550_unk	150 genes	Deletion	7,901,553		Behavioral abnormality, global development delay, microcephaly	(Miller et al., 2010)
nssv13646178_unk	2,161 genes	Duplication	135,285,622		Development delay and/or other significant development or morphological phenotypes	(Miller et al., 2010)
nssv13638976_unk, nssv13640749_unk	2,161 genes	Duplication	135,327,117		Abnormal facial shape, Intrauterine growth retardation, Micrognathia, Syndactyly, Ventricular septal defect	(Miller et al., 2010)
nssv13655969_unk	122 genes	Deletion	6,302,504		Developmental delay and/or other significant developmental or morphological phenotypes	(Miller et al., 2010)
nssv13655409_unk	928 genes	Deletion	53,839,386		Global developmental delay	(Miller et al., 2010)
nssv13653237_unk	717 genes	Duplication	42,143,651		Patent ductus arteriosus	(Miller et al., 2010)
nssv577306_dnovo	109 genes	Deletion	5,128,423		Dilatation, Hydronephrosis	(Kaminsky et al., 2011)
nssv577307_dnovo	176 genes	Deletion	8,175,579		Abnormal facial shape	(Kaminsky et al., 2011)
nssv1494941_unk	54 genes	Deletion	2,827,219		Autism, Failure to thrive	(Miller et al., 2010)
Benign						
nssv581618_unk	CYP2C18, CYP2C19	Deletion	158,935		Intellectual disability	(Miller et al., 2010)
nssv1608783_unk	CYP2C19	Deletion	20,211		Developmental delay and/or other significant developmental or morphological phenotypes	(Miller et al., 2010)
Likely Benign						
nssv13650654_unk	CYP2C18, CYP2C19	Deletion	121,822		Developmental delay and/or other significant developmental or morphological phenotypes	(Miller et al., 2010)
nssv13655256_unk	CYP2C18	Deletion	114,674		Developmental delay and/or other significant developmental or morphological phenotypes	(Miller et al., 2010)
nssv13647268_unk	CYP2C19, CYP2C9	Duplication	180,257		Developmental delay and/or other significant developmental or morphological phenotypes	(Miller et al., 2010)
nssv13650288_unk	CYP2C18	Deletion	48,162		Developmental delay and/or other significant developmental or morphological phenotypes	(Miller et al., 2010)
nssv585041_unk	CYP2C18	Deletion	112,487		Developmental delay and/or other significant developmental or morphological phenotypes	(Miller et al., 2010)

| CNV ID	Genes	CNV type	Region size	Frequency	Study population	Ref.
nssv13655998_unk	CYP2C8	Duplication	8,867		Developmental delay and/or other significant developmental or morphological phenotypes	(Miller et al., 2010)
Uncertain						
nssv13650363_unk	CYP2C18, CYP2C19	Deletion	240,911		Seizures	(Miller et al., 2010)
nssv583937_pat	PLCE1, NOC3L, TBC1D12, HELLS, CYP2C18, CYP2C19, CYP2C9, CYP2C8, PDLIM1, ACSM6	Deletion	969,724		Low-set ears	(Miller et al., 2010)
nssv580748_unk	CYP2C9, CYP2C8, PDLIM1, SORBS1	Duplication	473,822		Global developmental delay	(Kaminsky et al., 2011)
nssv1495397_unk	TBC1D12, HELLS, CYP2C18, CYP2C19, CYP2C9, CYP2C8, PDLIM1, SORBS1, ACSM6	Duplication	1,043,272		Autistic behavior, Global developmental delay	(Miller et al., 2010)
nssv3395040_unk	TBC1D12, HELLS, CYP2C18, CYP2C19, CYP2C9, CYP2C8, PDLIM1, SORBS1, ACSM6	Duplication	1,089,626		Seizures	(Miller et al., 2010)
DECIPHER						
184	175 genes	Deletion	12,434,019		Anterior creases of earlobe, capillary hemangiomas, hypoglycemia, intellectual disability, microcephaly	(Firth et al., 2009)
341717	122 genes	Duplication	8,517,656		Emotional lability, growth delay, intellectual disability	(Firth et al., 2009)
2578	232 genes	Duplication	17,187,727		Behavioral abnormality, constipation, deeply set eye, delayed speech and language development, intellectual disability, macrocephaly, pectus excavatum, plagiocephaly, short stature, ventricular septal defect	(Firth et al., 2009)
337109	PLCE1, NOC3L, TBC1D12, HELLS, CYP2C18, CYP2C19, CYP2C9, CYP2C8, PDLIM1	Duplication	1,080,102		Intellectual disability - moderate, oromotor apraxia, severe expressive language delay	(Firth et al., 2009)
292398	PLCE1, NOC3L, TBC1D12, HELLS, CYP2C18, CYP2C19, CYP2C9, CYP2C8, PDLIM1	Duplication	1,080,102		Anxiety, cognitive impairment	(Firth et al., 2009)
265438	PLCE1, NOC3L, TBC1D12, HELLS, CYP2C18, CYP2C19, CYP2C9, CYP2C8, PDLIM1	Duplication	1,080,102		-	(Firth et al., 2009)
260874	PLCE1, NOC3L, TBC1D12, HELLS, CYP2C18, CYP2C19, CYP2C9, CYP2C8, PDLIM1	Duplication	1,080,102		Delayed speech and language development, incomprehensible speech, intellectual disability - moderate, oromotor apraxia	(Firth et al., 2009)
318690	CYP2C18, CYP2C19	Deletion	120,242		Delayed speech and language development, EEG abnormality, moderate global developmental delay	(Firth et al., 2009)
300098	CYP2C18, CYP2C19	Deletion	158,935		Cognitive impairment	(Firth et al., 2009)
283724	CYP2C18, CYP2C19	Deletion	140,797		Abnormal facial shape, intellectual disability	(Firth et al., 2009)
278463	CYP2C19	Deletion	61,682		Behavioral abnormality	(Firth et al., 2009)
270359	CYP2C19, CYP2C8, CYP2C9	Duplication	288,534		Hypospadias, rudimentary fibula	(Firth et al., 2009)
256495	34 genes (CYP2C18 is not included)	Deletion	2,093,635		-	(Firth et al., 2009)
305972	CYP2C19, CYP2C8, CYP2C9	Duplication	344,191		-	(Firth et al., 2009)
272275	CYP2C8, CYP2C9, PDLIM1, SORBS1	Deletion	694,656		-	(Firth et al., 2009)

REFERENCES

Bush WS, Crosslin DR, Owusu-Obeng A, Wallace J, Almoguera B, Basford MA, … Ritchie MD (2016). Genetic variation among 82 pharmacogenes: The PGRNseq data from the eMERGE network. Clin Pharmacol Ther, 100(2), 160–169. doi:10.1002/cpt.350 [PubMed: 26857349]

Cahan P, Li Y, Izumi M, & Graubert TA (2009). The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells. Nat Genet, 41(4), 430–437. doi: 10.1038/ng.350 [PubMed: 19270704]

Carvalho CM, & Lupski JR (2016). Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet, 17(4), 224–238. doi:10.1038/nrg.2015.25 [PubMed: 26924765]

Caudle KE, Rettie AE, Whirl-Carrillo M, Smith LH, Mintzer S, Lee MT, … Clinical Pharmacogenetics Implementation, C. (2014). Clinical pharmacogenetics implementation consortium guidelines for CYP2C9 and HLA-B genotypes and phenytoin dosing. Clin Pharmacol Ther, 96(5), 542–548. doi:10.1038/clpt.2014.159 [PubMed: 25099164]

Cho HJ, Eom HS, Kim HJ, Kim IS, Lee GW, & Kong SY (2010). Glutathione-S-transferase genotypes influence the risk of chemotherapy-related toxicities and prognosis in Korean patients with diffuse large B-cell lymphoma. Cancer Genet Cytogenet, 198(1), 40–46. doi:10.1016/j.cancergencyto. 2009.12.004 [PubMed: 20303013]

Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, … Eichler EE (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet, 46(10), 1063–1071. doi:10.1038/ng.3092 [PubMed: 25217958]

Cook EH Jr., & Scherer SW (2008). Copy-number variations associated with neuropsychiatric conditions. Nature, 455(7215), 919–923. doi:10.1038/nature07458 [PubMed: 18923514]

Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, … Eichler EE (2011). A copy number variation morbidity map of developmental delay. Nat Genet, 43(9), 838–846. doi: 10.1038/ng.909 [PubMed: 21841781]

Douglas J, Tatton-Brown K, Coleman K, Guerrero S, Berg J, Cole TR, … Rahman N (2005). Partial NSD1 deletions cause 5% of Sotos syndrome and are readily identifiable by multiplex ligation dependent probe amplification. J Med Genet, 42(9), e56. doi:10.1136/jmg.2005.031930 [PubMed: 16140999]

Escaramis G, Docampo E, & Rabionet R (2015). A decade of structural variants: description, history and methods to detect structural variation. Brief Funct Genomics, 14(5), 305–314. doi:10.1093/ bfgp/elv014 [PubMed: 25877305]

Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, … Carter NP (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet, 84(4), 524–533. doi:10.1016/j.ajhg.2009.03.010 [PubMed: 19344873]

Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, & den Dunnen JT (2011). LOVD v.2.0: the next generation in gene variant databases. Hum Mutat, 32(5), 557–563. doi:10.1002/humu. 21438 [PubMed: 21520333]

Gaedigk A, Gaedigk R, & Leeder JS (2010). UGT2B17 and SULT1A1 gene copy number variation (CNV) detection by LabChip microfluidic technology. Clin Chem Lab Med, 48(5), 627–633. doi: 10.1515/CCLM.2010.128 [PubMed: 20192879]

Gaedigk A, Ingelman-Sundberg M, Miller NA, Leeder JS, Whirl-Carrillo M, Klein TE, & PharmVar Steering C (2018). The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. Clin Pharmacol Ther, 103(3), 399–401. doi:10.1002/cpt.910 [PubMed: 29134625]

Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Twist GP, Klein TE, Miller NA, & PharmVar Steering C (2019). The Evolution of PharmVar. Clin Pharmacol Ther, 105(1), 29–32. doi:10.1002/cpt.1275 [PubMed: 30536702]

Gaedigk A, Twist GP, & Leeder JS (2012). CYP2D6, SULT1A1 and UGT2B17 copy number variation: quantitative detection by multiplex PCR. Pharmacogenomics, 13(1), 91–111. doi: 10.2217/pgs.11.135 [PubMed: 22111604]

Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, … McVean GA (2010). A map of human genome variation from population-scale sequencing. Nature, 467(7319), 1061–1073. doi:10.1038/nature09534 [PubMed: 20981092]

Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, … McVean GA (2012). An integrated map of genetic variation from 1,092 human genomes. Nature, 491(7422), 56–65. doi:10.1038/nature11632 [PubMed: 23128226]

Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, … Abecasis GR (2015). A global reference for human genetic variation. Nature, 526(7571), 68–74. doi:10.1038/nature15393 [PubMed: 26432245]

Gjerde J, Hauglid M, Breilid H, Lundgren S, Varhaug JE, Kisanga ER, … Lien EA (2008). Effects of CYP2D6 and SULT1A1 genotypes including SULT1A1 gene copy number on tamoxifen metabolism. Ann Oncol, 19(1), 56–61. doi:10.1093/annonc/mdm434 [PubMed: 17947222]

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, … Ahuja SK (2005). The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science, 307(5714), 1434–1440. doi:10.1126/science.1101160 [PubMed: 15637236]

Gordon AS, Tabor HK, Johnson AD, Snively BM, Assimes TL, Auer PL, … Project, N. G. E. S. (2014). Quantifying rare, deleterious variation in 12 human cytochrome P450 drug-metabolism genes in a large-scale exome dataset. Hum Mol Genet, 23(8), 1957–1963. doi:10.1093/hmg/ddt588 [PubMed: 24282029]

Han Y, Jin X, Li H, Wang K, Gao J, Song L, & Lv Y (2017). Microarray analysis of copy-number variations and gene expression profiles in prostate cancer. Medicine (Baltimore), 96(28), e7264. doi:10.1097/MD.0000000000007264 [PubMed: 28700469]

He Y, Hoskins JM, & McLeod HL (2011). Copy number variants in pharmacogenetic genes. Trends Mol Med, 17(5), 244–251. doi:10.1016/j.molmed.2011.01.007 [PubMed: 21388883]

Henrichsen CN, Chaignat E, & Reymond A (2009). Copy number variants, diseases and gene expression. Hum Mol Genet, 18(R1), R1–8. doi:10.1093/hmg/ddp011 [PubMed: 19297395]

Hicks JK, Bishop JR, Sangkuhl K, Muller DJ, Ji Y, Leckband SG, … Clinical Pharmacogenetics Implementation, C. (2015). Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and CYP2C19 Genotypes and Dosing of Selective Serotonin Reuptake Inhibitors. Clin Pharmacol Ther, 98(2), 127–134. doi:10.1002/cpt.147 [PubMed: 25974703]

Hicks JK, Sangkuhl K, Swen JJ, Ellingrod VL, Muller DJ, Shimoda K, … Stingl JC (2017). Clinical pharmacogenetics implementation consortium guideline (CPIC) for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants: 2016 update. Clin Pharmacol Ther, 102(1), 37–44. doi:10.1002/cpt.597 [PubMed: 27997040]

Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, … Eichler EE (2009). Population analysis of large copy number variants and hotspots of human genetic disease. Am J Hum Genet, 84(2), 148–161. doi:10.1016/j.ajhg.2008.12.014 [PubMed: 19166990]

Johansson AC, & Feuk L (2011). Characterization of copy number-stable regions in the human genome. Hum Mutat, 32(8), 947–955. doi:10.1002/humu.21524 [PubMed: 21542059]

Johnson JA, Caudle KE, Gong L, Whirl-Carrillo M, Stein CM, Scott SA, … Wadelius M (2017). Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Pharmacogenetics-Guided Warfarin Dosing: 2017 Update. Clin Pharmacol Ther, 102(3), 397–404. doi:10.1002/cpt.668 [PubMed: 28198005]

Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, … Martin CL (2011). An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med, 13(9), 777–784. doi:10.1097/GIM.0b013e31822c79f9 [PubMed: 21844811]

Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, … Eichler EE (2008). Mapping and sequencing of structural variation from eight human genomes. Nature, 453(7191), 56–64. doi:10.1038/nature06862 [PubMed: 18451855]

Li J, Lao X, Zhang C, Tian L, Lu D, & Xu S (2014). Increased genetic diversity of ADME genes in African Americans compared with their putative ancestral source populations and implications for pharmacogenomics. BMC Genet, 15, 52. doi:10.1186/1471-2156-15-52 [PubMed: 24884825]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Liu P, Carvalho CM, Hastings PJ, & Lupski JR (2012). Mechanisms for recurrent and complex human genomic rearrangements. Curr Opin Genet Dev, 22(3), 211–220. doi:10.1016/j.gde.2012.02.012 [PubMed: 22440479]

Martis S, Mei H, Vijzelaar R, Edelmann L, Desnick RJ, & Scott SA (2013). Multi-ethnic cytochrome-P450 copy number profiling: novel pharmacogenetic alleles and mechanism of copy number variation formation. Pharmacogenomics J, 13(6), 558–566. doi:10.1038/tpj.2012.48 [PubMed: 23164804]

Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, … Ledbetter DH (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet, 86(5), 749–764. doi:10.1016/j.ajhg.2010.04.006 [PubMed: 20466091]

Moriyama B, Obeng AO, Barbarino J, Penzak SR, Henning SA, Scott SA, … Walsh TJ (2017). Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for CYP2C19 and Voriconazole Therapy. Clin Pharmacol Ther, 102(1), 45–51. doi:10.1002/cpt.583 [PubMed: 27981572]

Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, … Mooser V (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science, 337(6090), 100–104. doi:10.1126/science.1217876 [PubMed: 22604722]

Nishioka K, Hayashi S, Farrer MJ, Singleton AB, Yoshino H, Imai H, … Hattori N (2006). Clinical heterogeneity of alpha-synuclein gene duplication in Parkinson's disease. Ann Neurol, 59(2), 298–309. doi:10.1002/ana.20753 [PubMed: 16358335]

Reiner J, Karger L, Cohen N, Mehta L, Edelmann L, & Scott SA (2017). Chromosomal Microarray Detection of Constitutional Copy Number Variation Using Saliva DNA. J Mol Diagn, 19(3), 397–403. doi:10.1016/j.jmoldx.2016.11.006 [PubMed: 28315673]

Roa BB, Garcia CA, & Lupski JR (1991). Charcot-Marie-Tooth disease type 1A: molecular mechanisms of gene dosage and point mutation underlying a common inherited peripheral neuropathy. Int J Neurol, 25–26, 97–107 Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11980069

Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerriere A, Vital A, … Campion D (2006). APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet, 38(1), 24–26. doi:10.1038/ng1718 [PubMed: 16369530]

Ruderfer DM, Hamamsy T, Lek M, Karczewski KJ, Kavanagh D, Samocha KE, … Purcell SM (2016). Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. Nat Genet, 48(10), 1107–1111. doi:10.1038/ng.3638 [PubMed: 27533299]

Santos M, Niemi M, Hiratsuka M, Kumondai M, Ingelman-Sundberg M, Lauschke VM, & Rodriguez-Antona C (2018). Novel copy-number variations in pharmacogenes contribute to interindividual differences in drug pharmacokinetics. Genet Med, 20(6), 622–629. doi:10.1038/gim.2017.156 [PubMed: 29261188]

Schulze J, Johansson M, Thorngren JO, Garle M, Rane A, & Ekstrom L (2013). SULT2A1 Gene Copy Number Variation is Associated with Urinary Excretion Rate of Steroid Sulfates. Front Endocrinol (Lausanne), 4, 88. doi:10.3389/fendo.2013.00088 [PubMed: 23874324]

Scott SA, Cohen N, Brandt T, Toruner G, Desnick RJ, & Edelmann L (2010). Detection of low-level mosaicism and placental mosaicism by oligonucleotide array comparative genomic hybridization. Genet Med, 12(2), 85–92. doi:10.1097/GIM.0b013e3181cc75d0 [PubMed: 20084009]

Scott SA, Sangkuhl K, Stein CM, Hulot JS, Mega JL, Roden DM, … Clinical Pharmacogenetics Implementation, C. (2013). Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. Clin Pharmacol Ther, 94(3), 317–323. doi:10.1038/clpt.2013.105 [PubMed: 23698643]

Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, … Wigler M (2007). Strong association of de novo copy number mutations with autism. Science, 316(5823), 445–449. doi:10.1126/science.1138659 [PubMed: 17363630]

Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K, … Hakonarson H (2009). High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome Res, 19(9), 1682–1690. doi:10.1101/gr.083501.108 [PubMed: 19592680]

Simon T, Becquemont L, Mary-Krause M, de Waziers I, Beaune P, Funck-Brentano C, & Jaillon P (2000). Combined glutathione-S-transferase M1 and T1 genetic polymorphism and tacrine hepatotoxicity. Clin Pharmacol Ther, 67(4), 432–437. doi:10.1067/mcp.2000.104944 [PubMed: 10801254]

South ST, Lee C, Lamb AN, Higgins AW, Kearney HM, Working Group for the American College of Medical, G., & Genomics Laboratory Quality Assurance, C. (2013). ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. Genet Med, 15(11), 901–909. doi:10.1038/gim.2013.129 [PubMed: 24071793]

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, … Korbel JO (2015). An integrated map of structural variation in 2,504 human genomes. Nature, 526(7571), 75–81. doi: 10.1038/nature15394 [PubMed: 26432246]

Suktitipat B, Naktang C, Mhuantong W, Tularak T, Artiwet P, Pasomsap E, … Jinawath N (2014). Copy number variation in Thai population. PLoS One, 9(8), e104355. doi:10.1371/journal.pone. 0104355 [PubMed: 25118596]

Swen JJ, Nijenhuis M, de Boer A, Grandia L, Maitland-van der Zee AH, Mulder H, … Guchelaar HJ (2011). Pharmacogenetics: from bench to byte--an update of guidelines. Clin Pharmacol Ther, 89(5), 662–673. doi:10.1038/clpt.2011.34 [PubMed: 21412232]

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, … Project, N. E. S. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science, 337(6090), 64–69. doi:10.1126/science.1219240 [PubMed: 22604720]

Tervasmaki A, Winqvist R, Jukkola-Vuorinen A, & Pylkas K (2014). Recurrent CYP2C19 deletion allele is associated with triple-negative breast cancer. BMC Cancer, 14, 902. doi: 10.1186/1471-2407-14-902 [PubMed: 25466287]

Verma SS, Josyula N, Verma A, Zhang X, Veturi Y, Dewey FE, … Pendergrass SA (2018). Rare variants in drug target genes contributing to complex diseases, phenome-wide. Sci Rep, 8(1), 4624. doi:10.1038/s41598-018-22834-4 [PubMed: 29545597]

Vijzelaar R, Botton MR, Stolk L, Martis S, Desnick RJ, & Scott SA (2018). Multi-ethnic SULT1A1 copy number profiling with multiplex ligation-dependent probe amplification. Pharmacogenomics, 19(9), 761–770. doi:10.2217/pgs-2018-0047 [PubMed: 29790428]

Vogler C, Gschwind L, Rothlisberger B, Huber A, Filges I, Miny P, … Papassotiropoulos A (2010). Microarray-based maps of copy-number variant regions in European and sub-Saharan populations. PLoS One, 5(12), e15246. doi:10.1371/journal.pone.0015246 [PubMed: 21179565]

Wong LP, Ong RT, Poh WT, Liu X, Chen P, Li R, … Teo YY (2013). Deep whole-genome sequencing of 100 southeast Asian Malays. Am J Hum Genet, 92(1), 52–66. doi:10.1016/j.ajhg.2012.12.005 [PubMed: 23290073]
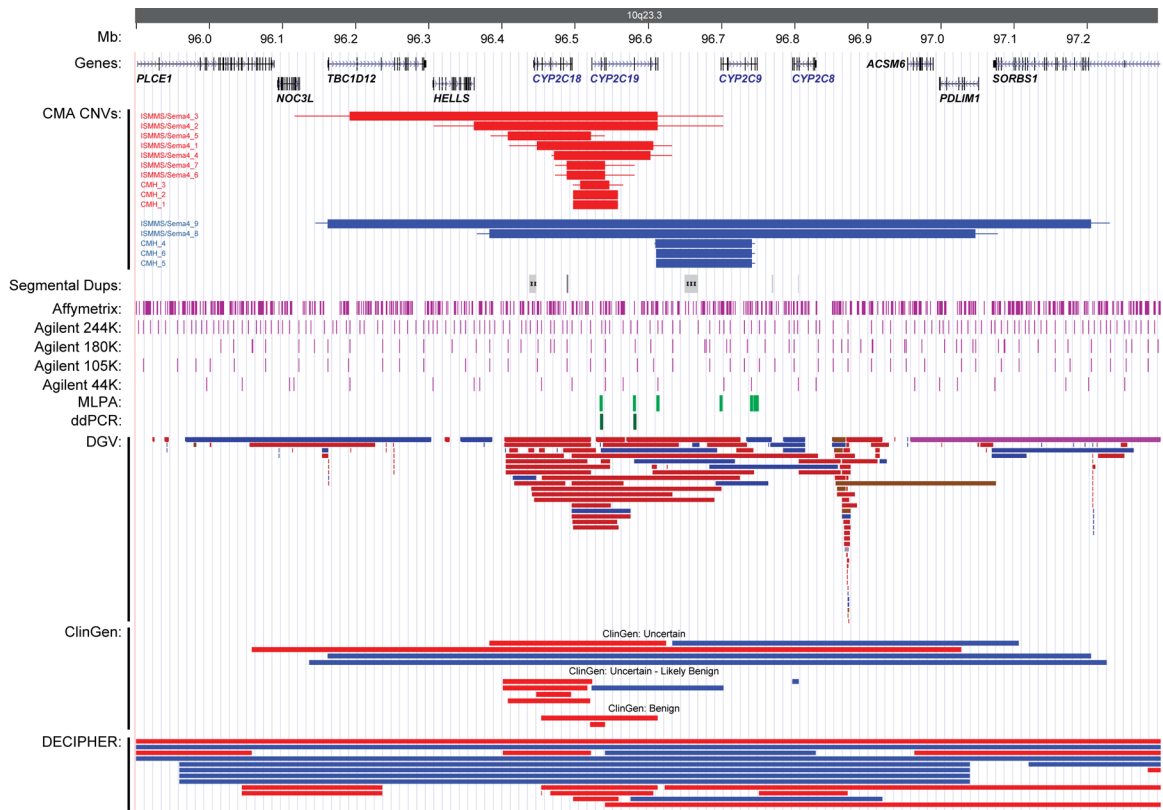
Figure 1.
Illustration of the chromosome 10q23.33 region with the identified *CYP2C* copy number variants (CNVs) noted in relation to the location of known human genes and transcripts from the UCSC Genes track, segmental duplications, microarray (purple), MLPA (light green), and ddPCR (dark green) probe locations, structural variants from the Database of Genomic Variants (DGV), ClinGen, and DECIPHER. Blue bars represent copy number gains (duplications), red bars represent copy number losses (deletions). The minimum and maximum CNV sizes in the clinical CMA cohorts are denoted by thick and thin horizontal bars, respectively.

Table 1.

CYP2C gene region copy number variants identified by clinical chromosomal microarray (CMA) testing.

Sample	ISCN 2016 Nomenclature	Microarray	Size (kb)	Genes Included	LOVD Variant ID
Discovery Cohort					
ISMMS/Sema4_6	arr[GRCh37] 10q23.33(96488443_96540495)x1	Agilent 180K	52.1	*CYP2C18* (exons 8–9), *CYP2C19* (exons 1–4)	0000484257
ISMMS/Sema4_7	arr[GRCh37] 10q23.33(96488443_96540495)x1	Agilent 180K	52.1	*CYP2C18* (exons 8–9), *CYP2C19* (exons 1–4)	0000484258
ISMMS/Sema4_5	arr[GRCh37] 10q23.33(96407740_96521658)x1	Agilent 180K	113.9	*CYP2C18* (exons 1–8), *CYP2C19* (exon 1)	0000484256
ISMMS/Sema4_4 [†]	arr[GRCh37] 10q23.33(96470997_96602860)x1	Agilent 180K	131.9	*CYP2C18* (exons 5–8), *CYP2C19* (exons 1–7)	0000484255
ISMMS/Sema4_1	arr[GRCh37] 10q23.33(96447479_96606715)x1	Agilent 105K	159.2	*CYP2C18* (exons 2–8), *CYP2C19* (exons 1–7)	0000484252
ISMMS/Sema4_2 [†]	arr[GRCh37] 10q23.33(96361629_96612764)x1	Agilent 44K	251.5	*HELLS, CYP2C18, CYP2C19*	0000484253
ISMMS/Sema4_3	arr[GRCh37] 10q23.33(96192063_96612764)x1	Agilent 44K	421.1	*TBC1D12, HELLS, CYP2C18, CYP2C19*	0000484254
ISMMS/Sema4_8 [†]	arr[GRCh37] 10q23.33q24.1(96383288_97047098)x3	Agilent 180K	663.8	*CYP2C18, CYP2C19, CYP2C9, CYP2C8, ACMS6, PDLIM1*	0000484259
ISMMS/Sema4_9	arr[GRCh37] 10q23.33q24.1(96161314_97204585)x3	Agilent 180K	1043.3	*CYP2C18, CYP2C19, CYP2C9, CYP2C8, ACMS6, PDLIM1, SORBS1* (exons 3–32)	0000484260
Replication Cohort					
CMH_3	arr[GRCh37] 10q23.33(96507212_96546422)x1	Agilent 244K	39.2	*CYP2C19* (exons 1–5)	0000484261
CMH_1	arr[GRCh37] 10q23.33(96497260_96558710)x1	Affymetrix	61.5	*CYP2C19* (exons 1–5)	0000484262
CMH_2	arr[GRCh37] 10q23.33(96497260_96558710)x1	Affymetrix	61.5	*CYP2C19* (exons 1–5)	0000484263
CMH_5	arr[GRCh37] 10q23.33(96610653_96741497)x3	Affymetrix	130.8	*CYP2C19* (exon 9), *CYP2C9* (exons 1–7)	0000484264
CMH_6	arr[GRCh37] 10q23.33(96610653_96741497)x3	Affymetrix	130.8	*CYP2C19* (exon 9), *CYP2C9* (exons 1–7)	0000484265
CMH_4	arr[GRCh37] 10q23.33(96609567_96741497)x3	Affymetrix	131.9	*CYP2C19* (exons 8–9), *CYP2C9* (exons 1–7)	0000484266

[†] Ethnicity for individuals ISMMS/Sema4_2, ISMMS/Sema4_4, and ISMMS/Sema4_8 is Hispanic/Latino, White, and Asian, respectively. Ethnicity was not available for all other reported subjects.

Affymetrix: refers to the Cytoscan® microarray; CMH: Children's Mercy Hospital; ISMMS: Icahn School of Medicine at Mount Sinai; LOVD: Leiden Open Variation Database (https://databases.lovd.nl).

Table 2.

CYP2C copy number variant (CNV) confirmation by multiplex ligation-dependent probe amplification (MLPA).

SAMPLE ID	*CYP2C19*			*CYP2C9*				
	Exon 2	Exon 6	Exon 9	Exon 1	Exon 7	Exon 8A	Exon 8B	Exon 9
ISMMS/Sema4_4	0.499	0.458	0.517	0.928	0.962	0.953	0.989	0.961
ISMMS/Sema4_5 *	0.977	1.030	1.004	1.016	0.995	0.982	0.949	1.075
ISMMS/Sema4_6	0.508	0.903	0.968	0.953	0.959	1.005	1.041	1.014
ISMMS/Sema4_7	0.481	0.854	0.992	1.004	1.010	1.027	1.064	0.984
ISMMS/Sema4_9	1.440	1.665	1.297	1.209	1.561	1.368	1.487	1.577

*
A 113.9 kb deletion was detected by CMA in this sample that included exons 1–8 of *CYP2C18* and only exon 1 of *CYP2C19*.

Light gray shaded cells indicate heterozygous deletion or duplication by CMA testing. MLPA ratios: one copy >0.25 and <0.75; two copies: >0.75 and <1.25; three copies >1.25 and <1.7.

Table 3.

CYP2C copy number variant (CNV) allele and carrier frequencies.

Cohort	CYP2C18			CYP2C19			CYP2C9			CYP2C8			Any CYP2C gene		
	Carriers/Cohort	Allele Frequency (%)	Carrier Frequency (%)	Carriers/Cohort	Allele Frequency (%)	Carrier Frequency (%)	Carriers/Cohort	Allele Frequency (%)	Carrier Frequency (%)	Carriers/Cohort	Allele Frequency (%)	Carrier Frequency (%)	Carriers/Cohort	Allele Frequency (%)	Carrier Frequency (%)
Deletion															
ISMMS/Sema4	7/11096	0.032	0.063	7/11096	0.032	0.063	0/11096	-	-	0/11096	-	-	7/11096	0.032	0.063
CMH	0/9760	-	-	3/9760	0.015	0.031	0/9760	-	-	0/9760	-	-	3/9760	0.015	0.031
DGV	38/41392	0.046	0.092	31/41392	0.037	0.075	4/41392	0.005	0.010	3/41392	0.004	0.007	50/41392	0.060	0.121
ClinGen	7/37447	0.009	0.019	5/37447	0.007	0.013	1/37447	0.001	0.003	1/37447	0.001	0.003	8/37447	0.011	0.021
Total:	52/99695	0.026	0.052	46/99695	0.023	0.046	5/99695	0.003	0.005	4/99695	0.002	0.004	68/99695	0.034	0.068
Duplication															
ISMMS/Sema4	2/11096	0.009	0.018	2/11096	0.009	0.018	2/11096	0.009	0.018	2/11096	0.009	0.018	2/11096	0.009	0.018
CMH	0/9760	-	-	3/9760	0.015	0.031	3/9760	0.015	0.031	0/9760	-	-	3/9760	0.015	0.031
DGV	1/41392	0.001	0.002	3/41392	0.004	0.007	4/41392	0.005	0.010	4/41392	0.005	0.010	9/41392	0.016	0.022
ClinGen	2/37447	0.003	0.005	3/37447	0.004	0.003	4/37447	0.005	0.011	4/37447	0.005	0.011	5/37447	0.007	0.013
Total:	5/99695	0.002	0.005	10/99695	0.005	0.010	12/99695	0.006	0.012	10/99695	0.005	0.010	18/99695	0.009	0.018
Deletion/Duplication															
ISMMS/Sema4	9/11096	0.041	0.081	9/11096	0.041	0.081	2/11096	0.009	0.018	2/11096	0.009	0.018	9/11096	0.041	0.081
CMH	0/9760	-	-	6/9760	0.031	0.061	3/9760	0.015	0.031	0/9760	-	-	6/9760	0.031	0.061
DGV	39/41392	0.047	0.094	34/41392	0.041	0.082	8/41392	0.010	0.019	7/41392	0.008	0.017	59/41392	0.076	0.143
ClinGen	9/37447	0.012	0.024	8/37447	0.011	0.021	5/37447	0.007	0.013	5/37447	0.007	0.013	13/37447	0.016	0.033
Total:	57/99695	0.029	0.057	57/99695	0.029	0.057	18/99695	0.009	0.018	14/99695	0.007	0.014	87/99695	0.044	0.087