

## **Drug Metabolism Disposition**

***The CYP2C19 intron 2 branch point SNP is the ancestral polymorphism contributing to the poor metabolizer phenotype in livers with CYP2C19\*35 and CYP2C19\*2 alleles***

Amarjit S. Chaudhry, Bhagwat Prasad, Yoshiyuki Shirasaka, Alison Fohner, David Finkelstein, Yiping Fan, Shuoguo Wang, Gang Wu, Eleni Aklillu, Sarah Sim, Kenneth E. Thummel and Erin G. Schuetz.

**Supplemental Table 1: Primers used for PCR amplification and sequencing of *CYP2C19* exons from genomic DNA.**

Exons	FP/RP	Primer sequence	PCR amplified fragment size (base pairs)
Exon 1	FP	5'-AGTGGGCCTAGGTGATTGCCACTT-3'	410
	RP	5'-TCAAAGTATTTACTTACAATGATCTC-3'	
Exon 2-3	FP	5'-AAAATATGAATCTAAGTCAGGCTTAGT-3'	607
	RP	5'-GGAGAGCAGTCCAGAAAGGTCACTGATA-3'	
Exon 4	FP	5'-TGCTTTAACCGAATTCAAGG-3'	383
	RP	5'-AAAATGTACTTCAGGGCTTGG-3'	
Exon 5	FP	5'-CAACCAGAGCTTGGCATATTG-3'	409
	RP	5'-TGATGCTTACTGGATATTCACTGC-3'	
Exon 6	FP	5'-AAACTGGCACAAAGACAGGGATG-3'	456
	RP	5'-AAATTGGGACAGATTACAGCTGCG-3'	
Exon 7	FP	5'-AATTGCTAGAACAAATGTTCCATTTC-3'	327
	RP	5'-AGAGGGTAAGAACATCATACTGTGA-3'	
Exon 8	FP	5'-CCACTGTTCTAACACCTTCGTGA-3'	284
	RP	5'-GAAGGCACATGTAAGTCCAAGTGA-3'	
Exon 9	FP	5'-ATCTACTCATCCCTCCTATGATTACCG-3'	529
	RP	5'-ATGTGGCACTCAATGTAAGTATTATAGA-3'	

**Supplemental Table 2: Optimized MS/MS parameters of CYP2C19 surrogate peptides used for protein quantification**

Peptide	Amino acid position	Parent ion	Product ion	Cone (V)	CE (eV)
60-73					
IYGPVFTLYFGLER	(exon 2)	838.2	998.3	52	28
		838.2	1145.4	52	28
384-399					
GTTILTSLTSVLHDNK	(exon 8)	567.3	664.4	35	17
		567.3	607.8	35	17
ASPC[160]DPTFILGC[160]AP	161-185				
C[160]NVIC[160]SIIFQK*	(exon 4)	1434.68	535.32	60	42
		1434.68	989.44	60	42

\*C[160] indicates S-carbamidomethylated cysteine after alkylation of native peptide.

**Supplemental Table 3: *CYP2C19*\*2/\*35 diplotypes and *CYP2C19*\*2 and *CYP2C19*\*35 allele frequencies in different populations (1000 genome-Phase 3 data)**

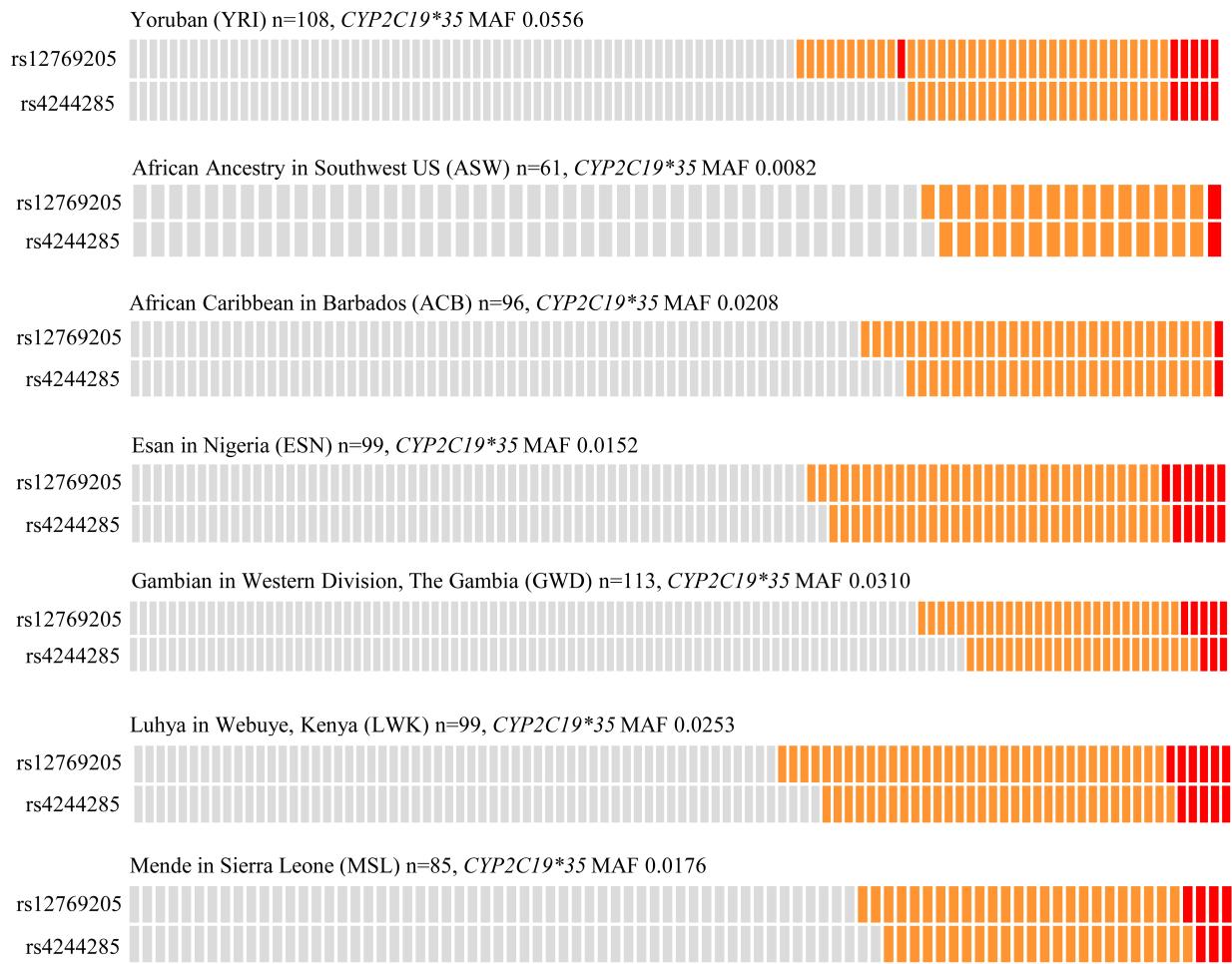
	All Africans*	Yorubans (YRI)	Caucasians (CEU)	Chinese (CHB)
Diplotype	n =661	n=108	n=99	n=103
<i>CYP2C19</i> *1/*1	430	66	75	41
<i>CYP2C19</i> *1/*2	174	26	22	55
<i>CYP2C19</i> *1/*35	28	10	0	0
<i>CYP2C19</i> *2/*35	5	0	0	0
<i>CYP2C19</i> *2/*2	23	5	2	7
<i>CYP2C19</i> *35/*35	1	1	0	0

Allele Frequency				
<i>CYP2C19</i> *1	0.8033	0.7778	0.8687	0.6650
<i>CYP2C19</i> *2	0.1702	0.1667	0.1313	0.3350
<i>CYP2C19</i> *35	0.0265	0.0556	0.0000	0.0000

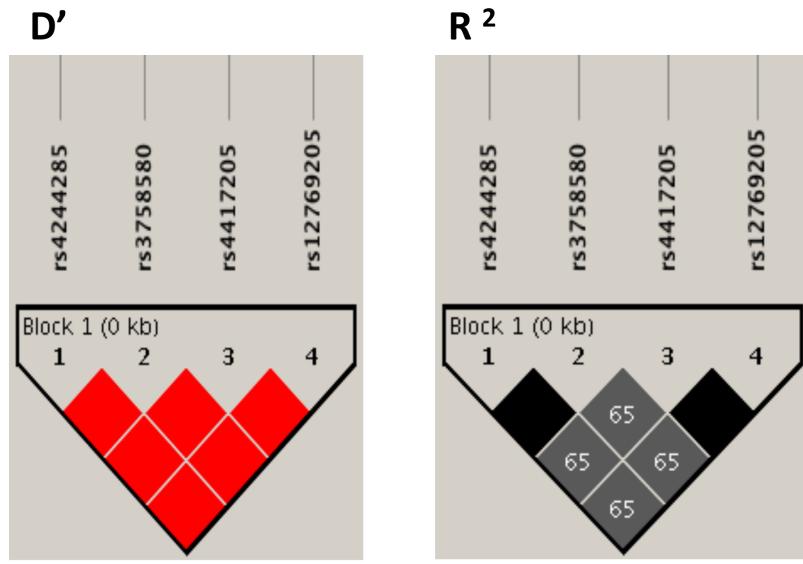
\*All Africans includes all seven populations of African descent (YRI, ASW, ACB, ESN, GWD, LWK and MSL) as defined in Supplemental Figure 1.

**Supplemental Table 4: Full length sequencing of the *CYP2C19\*35* cDNA**

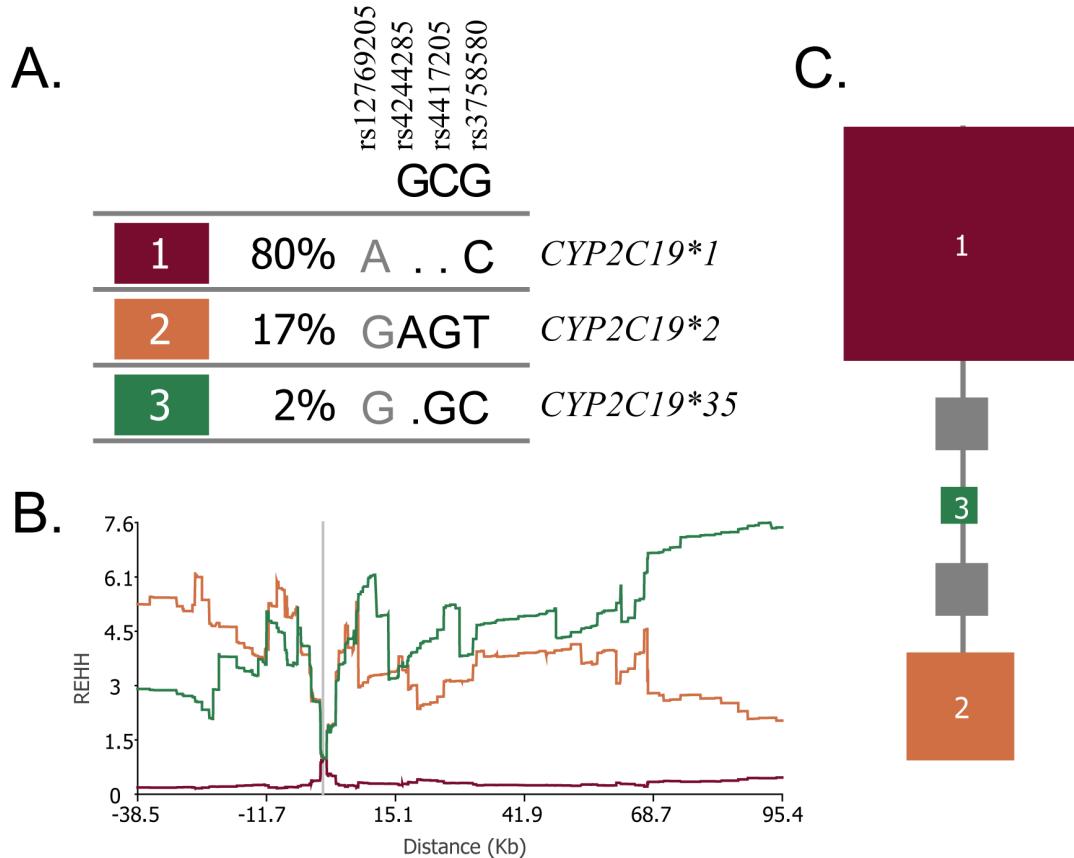
SNP rsID	Liver Genotype observed	CYP2C19*35 allele	Gene position	mRNA Position	Location in intron	Exon	Amino acid change
rs17885098	CT	T	99C>T	99C>T	-	Ex 1	Pro33Pro
rs12769205	AG	G	12662A>G	-	Ex 3-23A>G	Int 2	-
rs3758581	GG	G	80161A>G	991A>G	-	Ex 7	Ile331Val



**Supplemental Figure 1: Visual genotypes of rs12769205 and rs4244285 in populations of African descent (1000 genomes phase 3 data).** Visual genotypes for rs12769205 and rs4244285 in 1000 genomes samples (genotypes downloaded from <http://browser.1000genomes.org/index.html>). The populations of African descent are each listed with the number of individuals (n) and the *CYP2C19\*35* minor allele frequency. Grey, orange and red boxes indicate homozygous wild-type, heterozygous and homozygous variant genotypes, respectively.



**Supplemental Figure 2: Linkage disequilibrium (LD) map for the common *CYP2C19* allelic variants in the YRI population.** (Left panel) A LD map of the *CYP2C19* SNPs rs4244285, rs3758580, rs4417205 and rs12769205 in the YRI population was created using Haploview 4.2. The red squares show complete LD with statistical significance ( $|D'| = 1$ , LOD  $> 2$ ). (Right panel)  $R^2$  LD values. The black squares show the SNPs have the highest correlation with each other, the gray boxes indicate a correlation of 0.65. The LD values in the figure are scaled from 1.0 to 100 for visual clarity.



**Supplemental Figure 3: *CYP2C19* haplotype frequencies, extended haplotype homozygosity, and ancestral tree in other African populations in the 1000 genome project.**

The African populations (excluding YRI) living on the African continent (LWK, Luhya in Webuye, Kenya; GWD, Gambian in Western divisions in the Gambia; MSL, Mende in Sierra Leone; and ESN, Esan in Nigeria) were combined ( $n=396$ ). (A) Sweep was used to determine *CYP2C19* haplotypes (SNP positions rs12769205, rs4244285, rs4417205 and rs3758580) and their frequencies were determined (see Fig. legend 9). (B) Relative extended haplotype homozygosity (REHH) for each *CYP2C19* haplotype with the core of the haplotypes centered on rs4244285. Both haplotypes containing SNP rs12769205 either alone (haplotype 3 (green, *CYP2C19\*35*)), or with rs4244285 (haplotype 2 (orange, *CYP2C19\*2*)) show extended haplotype homozygosity REHH. (C) Phylogenetic tree of the *CYP2C19* haplotypes. Haplotypes closer to the ancestral are at the top of the figure. The area of the squares is proportional to the frequency of the haplotype. The gray squares represent haplotypes not present in the data, but that are missing links in the phylogeny. The program determined the ancestral root of the tree was *CYP2C19\*1*.