SUPPLEMENTAL MATERIAL

Supplemental Table S1. Allele frequencies for 11 *CYP2E1* polymorphisms in 50 populations. Frequencies are given for the ancestral allele for the SNPs. The most common allele is shown for the VNTR since the ancestral allele cannot be determined.

Supplemental Table S2. Pairwise LD values measured as r^2 averaged across the 50 populations studied.

Supplemental Table S3. The 16 commonly occurring haplotypes explained as combinations of haplotypes at three core regions defined by markers 1-5, markers 6-9, and markers 10-11. The numbers refer to the haplotypes in Figure 3 for each of the cores. Each of these core haplotypes in Figure 3 can be explained as simple accumulation of mutations with no internal recombination.

Supplemental Figure S1. The Fst values and average heterozygosities plotted by physical position along the chromosome for each of the 11 markers in *CYP2E1* region based on the 50 population samples.

<u>Supplemental Figure S2a</u>. HAPLOT diagram showing regions of high LD across 11 markers at *CYP2E1*, based on r^2 values. The default parameter values in HAPLOT were used.

<u>Supplemental Figure S2b</u>. HAPLOT diagram showing regions of high LD across 11 markers at *CYP2E1*, based on D′ values. The default parameter values in HAPLOT were used.

<u>Supplemental Figure S3</u>. Haplotype frequencies for core A (5 upstream markers studied: 1-5 in

Table 1) in the 50 population samples studied. The 5 common haplotypes with frequencies of 5% or more are shown. Rare haplotypes are combined in the residual class.

Supplemental Figure S4. Haplotype frequencies for core B (markers 6-9 in Table 1) in the 50 population samples studied. The 4 common hapltoypes with frequencies of 5% or more are shown. Rare haplotypes are combined in the residual class.

Supplemental Figure S5. Haplotype frequencies for core C (markers 10-11 in Table 1) in 50 population samples. Three of 4 haplotypes occur at common frequencies of 5% or more.

Supplemental Table S1.
Allele frequencies for 11 CYP2E1 markers in 50 population samples--ancestral allele for SNPs

Markers:	VNTR	rs3813867	rs2031920	rs2070672	rs6413420	rs6413419	rs915909	rs4646976	rs6413432	rs2070676	rs2515641
Alleles:	6rpt	G	C	A	G	G	C	A	A	G	C
Biaka	0.667	0.971	0.993	0.669	0.906	0.942	0.877	0.921	0.229	0.645	0.442
Mbuti	0.605	0.974	1.000	0.566	1.000	0.962	0.821	0.961	0.276	0.538	0.462
Yoruba	0.860	0.968	0.994	0.833	1.000	0.756	0.948	0.910	0.073	0.737	0.314
Ibo	0.917	0.948	1.000	0.917	0.990	0.667	0.927	0.979	0.062	0.649	0.375
Hausa	0.908	0.973	1.000	0.895	0.974	0.750	0.846	0.959	0.026	0.744	0.308
Chagga	0.867	0.933	1.000	0.889	0.978	0.844	0.889	0.922	0.056	0.811	0.233
Masai	0.941	1.000	1.000	0.950	1.000	0.825	0.925	1.000	0.050	0.700	0.325
Sandawe	0.859	1.000	0.974	0.862	0.900	0.912	0.962	1.000	0.012	0.628	0.362
AfrAmericans	0.878	0.918	0.978	0.896	0.989	0.802	0.906	0.933	0.074	0.566	0.456
Ethiopians	0.891	0.984	1.000	0.891	0.891	1.000	1.000	1.000	0.047	0.391	0.656
Somali	0.825	0.972	1.000	0.800	0.900	0.975	1.000	0.925	0.132	0.500	0.500
Yemenites	0.930	1.000	1.000	0.930	0.919	0.895	0.977	1.000	0.070	0.267	0.744
Druze	0.951	0.960	0.937	0.947	0.966	0.947	1.000	0.964	0.054	0.152	0.854
Samaritans	0.950	1.000	0.988	0.936	0.838	0.988	1.000	1.000	0.012	0.200	0.800
Ashkenazi	0.975	0.926	0.939	0.975	0.944	0.981	1.000	0.994	0.125	0.160	0.850
Adygei	0.943	0.991	0.991	0.935	0.944	0.907	1.000	0.981	0.113	0.170	0.833
Chuvash	0.976	0.964	0.961	0.976	0.929	0.988	1.000	0.986	0.071	0.119	0.881
Hungarians	0.983	0.983	0.977	0.983	0.961	0.955	1.000	0.989	0.062	0.152	0.848
Russians_A	0.970	0.969	0.970	0.970	0.955	0.955	1.000	1.000	0.076	0.121	0.879
Russians_V	0.957	0.979	0.990	0.958	0.969	0.958	1.000	0.979	0.085	0.117	0.885
Finns	0.986	0.972	0.986	0.986	0.944	0.972	1.000	0.986	0.186	0.129	0.875
Danes	0.990	1.000	1.000	1.000	0.951	0.990	1.000	1.000	0.147	0.088	0.912
Irish	0.996	0.965	0.961	0.996	0.969	0.978	1.000	0.991	0.112	0.062	0.939
EuroAmericans	0.984	0.951	0.934	0.984	0.929	0.967	1.000	1.000	0.124	0.122	0.864
Komi Zyriane	0.968	1.000	1.000	0.968	0.926	0.913	1.000	0.979	0.032	0.191	0.809
Khanty	0.980	0.950	0.947	0.980	1.000	1.000	1.000	0.980	0.100	0.030	0.970
Mohanna	0.798	1.000	1.000	0.798	0.904	1.000	0.990	0.970	0.183	0.093	0.913
Negroid Makrani	0.750	0.981	1.000	0.732	0.911	0.911	0.981	0.946	0.240	0.426	0.630
Hazara	0.931	0.931	0.931	0.931	0.929	1.000	1.000	0.897	0.155	0.172	0.821
Keralites	0.767	0.983	0.983	0.767	0.948	0.946	0.983	0.917	0.150	0.207	0.800
Yakut	0.892	0.912	0.912	0.892	1.000	1.000	1.000	0.890	0.108	0.078	0.922
Nasioi	1.000	0.978	0.977	1.000	1.000	1.000	1.000	1.000	0.000	0.022	1.000
Micronesians	0.944	0.824	0.838	0.932	0.973	0.986	1.000	0.878	0.095	0.189	0.824
Laotians	0.765	0.811	0.811	0.771	0.987	1.000	0.996	0.836	0.267	0.186	0.819
Cambodians	0.875	0.917	0.958	0.860	0.920	1.000	1.000	0.875	0.109	0.167	0.800
SF Chinese	0.750	0.776	0.750	0.759	1.000	1.000	1.000	0.758	0.269	0.224	0.788
TW Chinese	0.850	0.730	0.806	0.847	1.000	1.000	1.000	0.850	0.327	0.130	0.870
Hakka	0.707	0.817	0.817	0.700	1.000	1.000	1.000	0.732	0.221	0.268	0.707
Koreans	0.815	0.824	0.824	0.815	1.000	1.000	1.000	0.868	0.231	0.139	0.870
Japanese	0.784	0.706	0.735	0.786	1.000	0.990	1.000	0.850	0.392	0.147	0.847
Ami	0.738	0.923	1.000	0.738	1.000	1.000	1.000	0.812	0.100	0.256	0.762
Atayal	0.810	0.622	0.690	0.805	1.000	1.000	1.000	0.857	0.429	0.119	0.866
Cheyenne	0.786	0.936	0.929	0.769	1.000	1.000	1.000	0.817	0.098	0.223	0.773
Pima_AZ	0.873	0.784	0.780	0.870	1.000	1.000	1.000	0.873	0.220	0.127	0.873
Pima_MX	0.943	0.528	0.528	0.943	1.000	1.000	1.000	0.923	0.462	0.057	0.943
Maya	0.913	0.700	0.735	0.939	0.990	0.990	1.000	0.942	0.275	0.100	0.900
Quechua	0.957	0.804	0.804	0.957	1.000	1.000	1.000	0.957	0.196	0.043	0.957
Ticuna	0.854	0.698	0.692	0.854	1.000	1.000	1.000	0.856	0.315	0.162	0.854
R.Surui	0.978	0.696	0.711	0.978	1.000	1.000	1.000	0.978	0.318	0.022	0.978
Karitiana	0.954	0.528	0.591	0.955	1.000	1.000	1.000	0.956	0.455	0.045	0.955

Supplemental Table S2.

Pairwise LD values for the 11 markers measured as r2 averaged across population samples

		1	2	3	4	5	6	7	8	9	10	11
1	VNTR		48	25	48	48	45	21	47	47	48	48
2	rs3813867	0.25		25	48	49	45	21	47	48	49	49
3	rs2031920	0.14	0.02		25	25	24	11	25	25	25	25
4	rs2070672	0.81	0.39	0.13		48	45	21	47	47	48	48
5	rs6413420	0.31	0.60	0.05	0.41		45	21	47	48	49	49
6	rs6413419	0.40	0.04	0.02	0.35	0.17		21	44	44	45	45
7	rs915909	0.16	0.10	0.09	0.11	0.08	0.00		21	20	21	21
8	rs4646976	0.75	0.27	0.09	0.61	0.32	0.32	0.18		46	47	47
9	rs6413432	0.12	0.31	0.04	0.10	0.57	0.22	0.01	0.16		48	48
10	rs2070676	0.20	0.50	0.04	0.34	0.62	0.21	0.11	0.21	0.42		49
11	rs2515641	0.04	0.53	0.07	0.15	0.43	0.04	0.06	0.04	0.30	0.69	

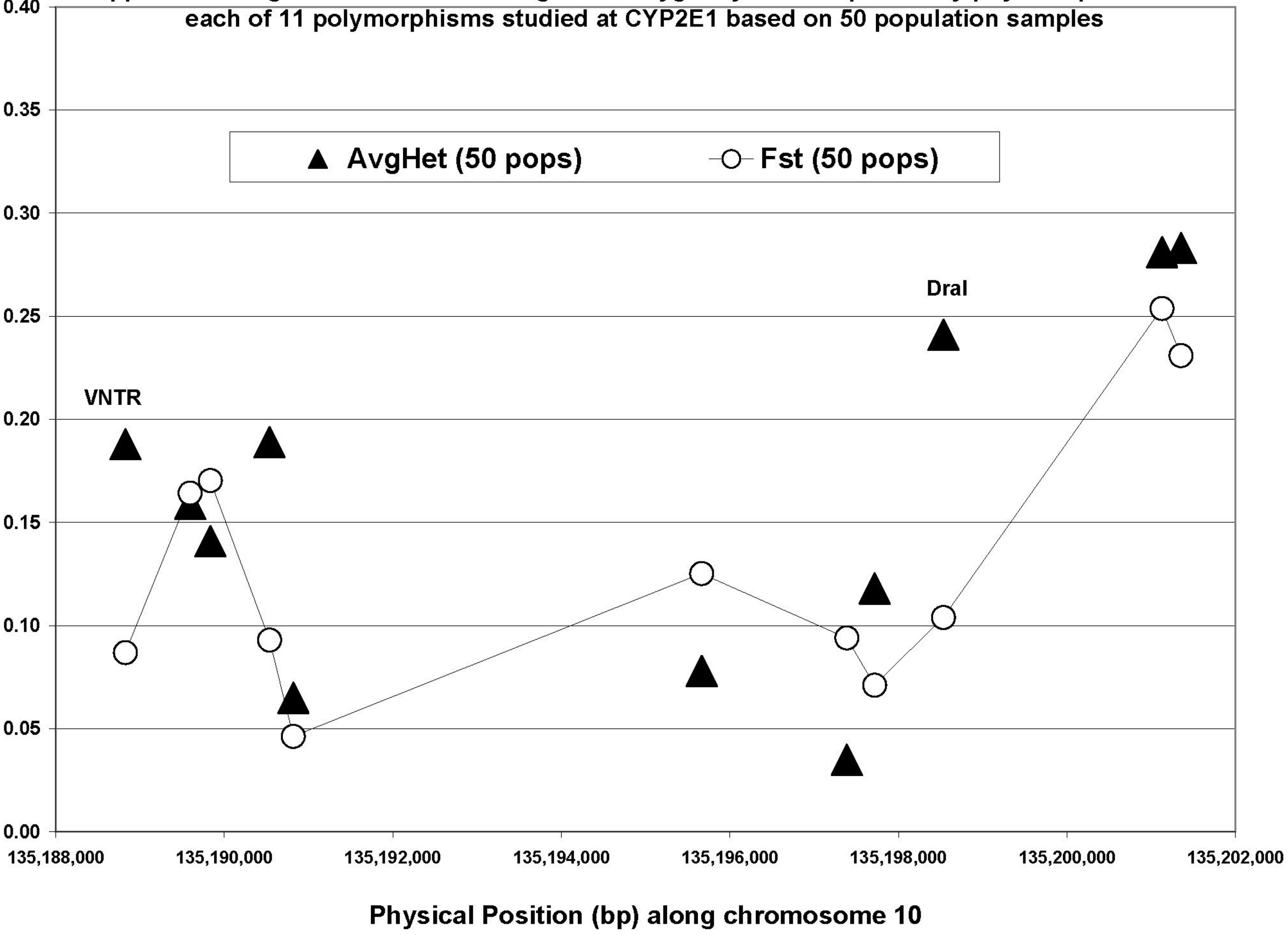
Note:

The lower triangle contains LD values and the upper triangle contains the number of populations contributing. All populations were typed for all the SNPs and the VNTR, but when a population sample is fixed for an allele (i.e. no variation for one or both of the paired polymorphisms), no LD value can be calculated.

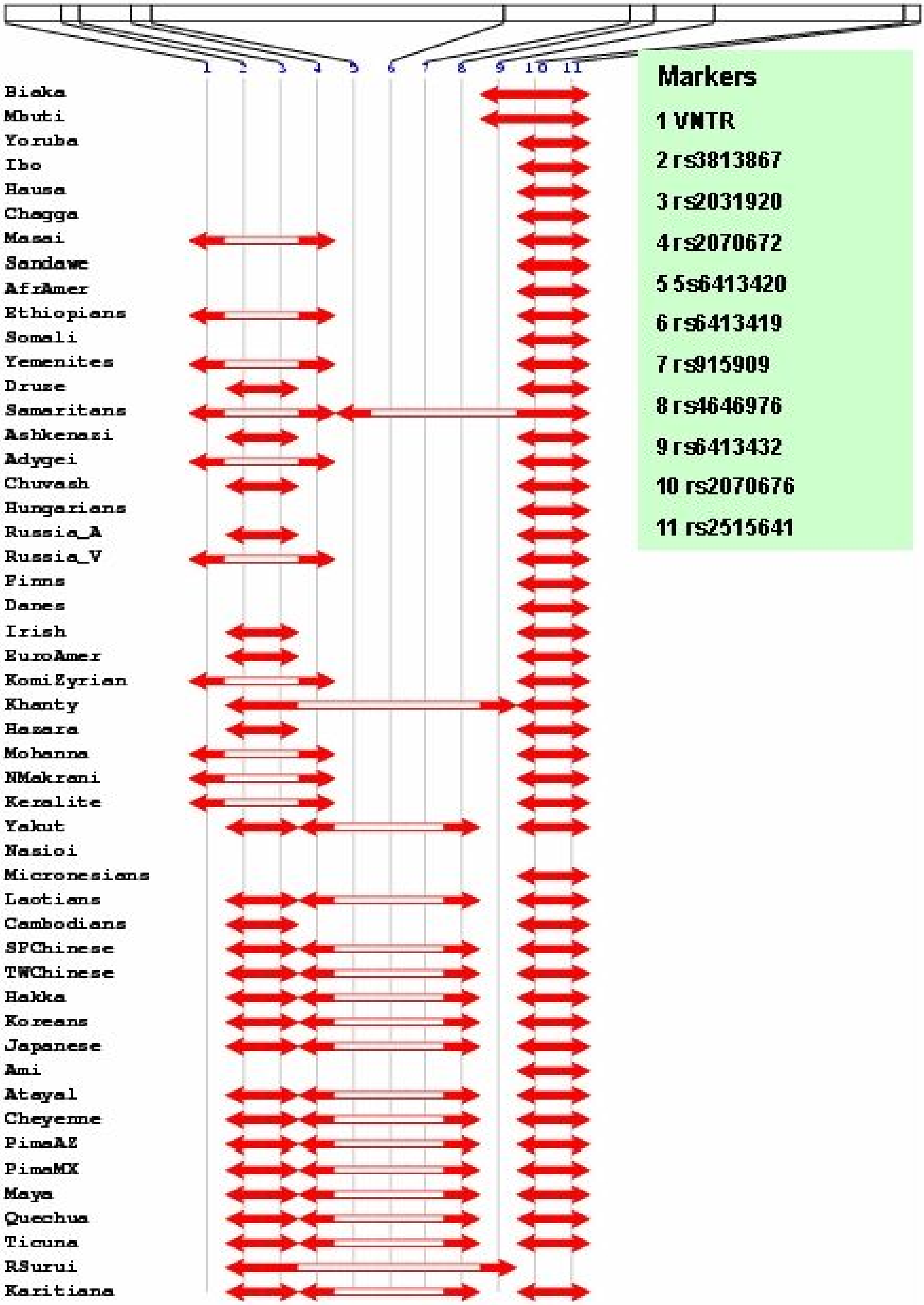
Supplemental Table S3. The 16 commonly occurring haplotypes explained as combinations of haplotypes at three core regions defined by markers 1-5, markers 6-9, and markers 10-11.

Haplotypes	Core A markers 1-5	Core B markers 6-9	Core C markers 10-11
Ancestral ?GCAGGCAAGC	1	1	1
a:6GCAGGCATCC	1	2	3
b:6GCAGGCATGT	1	2	2
c:6CTAGGCAACC	5	1	3
d:8GCGGGCGTGT	2	3	2
e:8GCGGGCAACC	2	1	3
f:6GCATGCATGT	3	2	2
g:6GCAGACATGT	1	4	2
h:6GCAGGCAACC	1	1	3
i:6GCAGGTATGT	1	5	2
j:8GCGGGCATGT	2	2	2
k:6GCAGACATCC	1	4	3
l:8GCGGGCATCC	2	2	3
m:6CCAGGCAACC	4	1	3
n:6CCAGACATCC	4	4	3
o:6GCATGCATCC	3	2	3
p:6CCAGGCGTGT	4	3	2

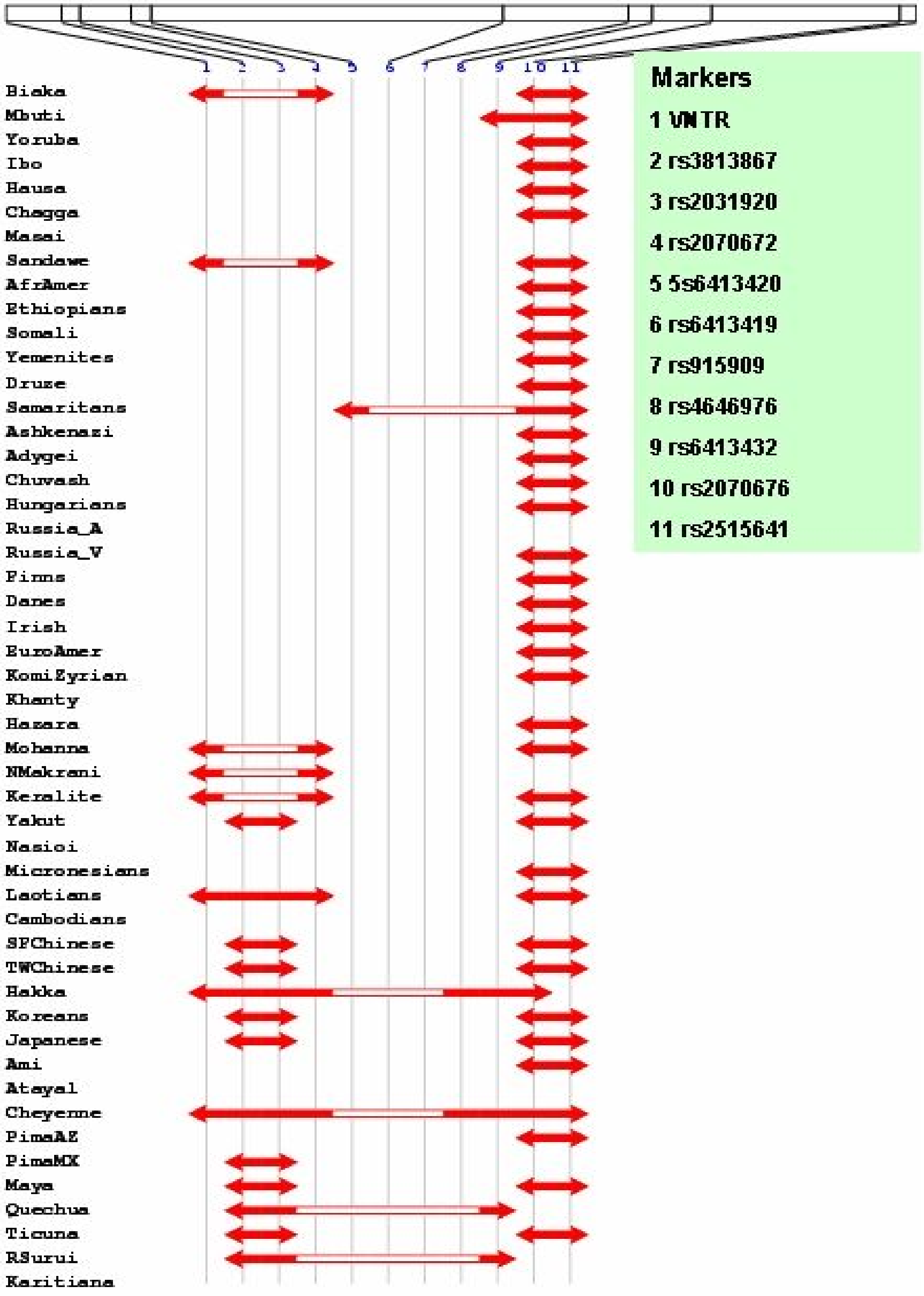
Supplemental Figure S1. Fst and average heterozygosity values plotted by physical position for each of 11 polymorphisms studied at CYP2E1 based on 50 population samples



Suppl. Fig S2a. Regions of high LD based on r2 values at *CYP2E1* (Figure generated by HAPLOT)

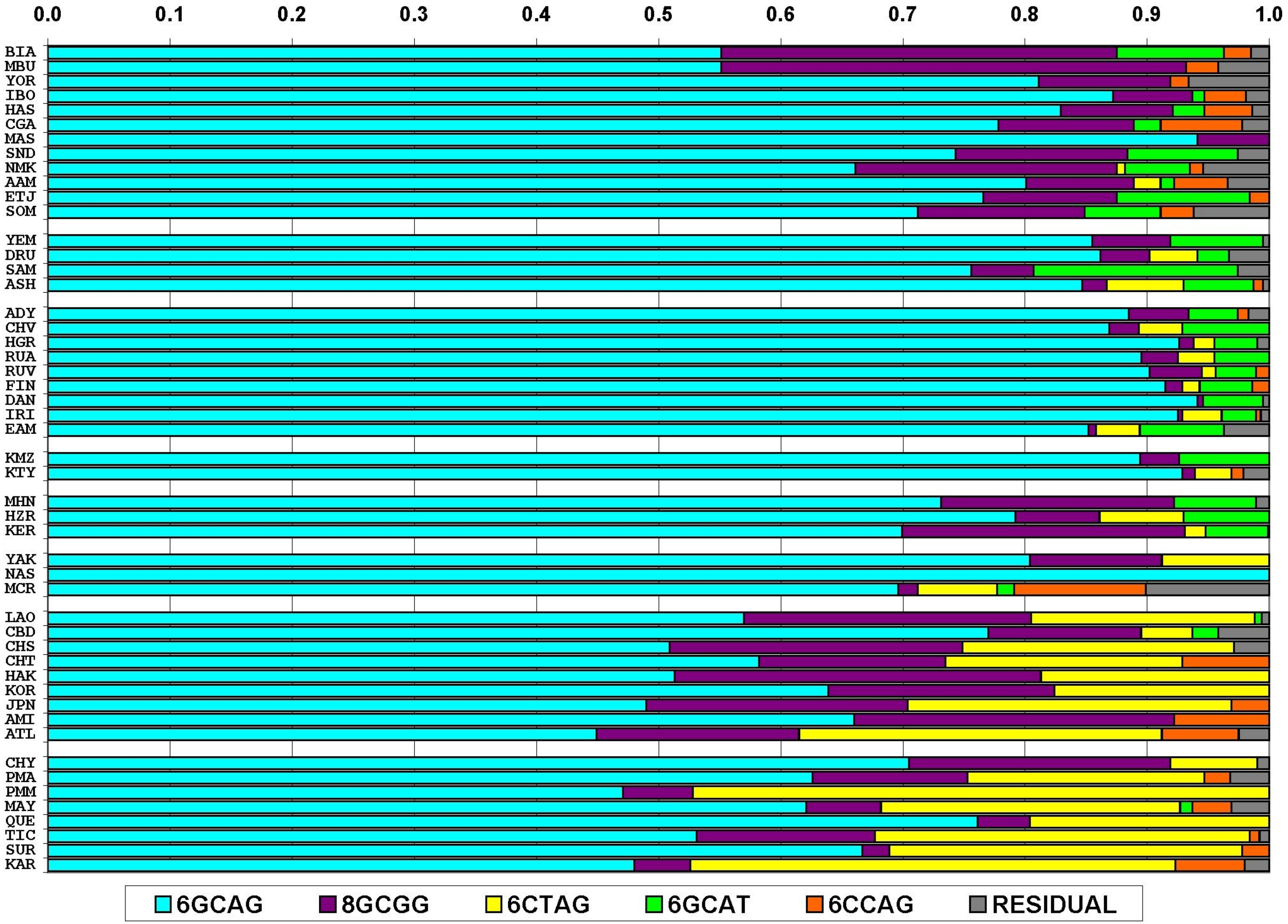


Suppl.Fig S2b. Regions of high LD based on D' values at *CYP2E1* (Figure generated by HAPLOT)



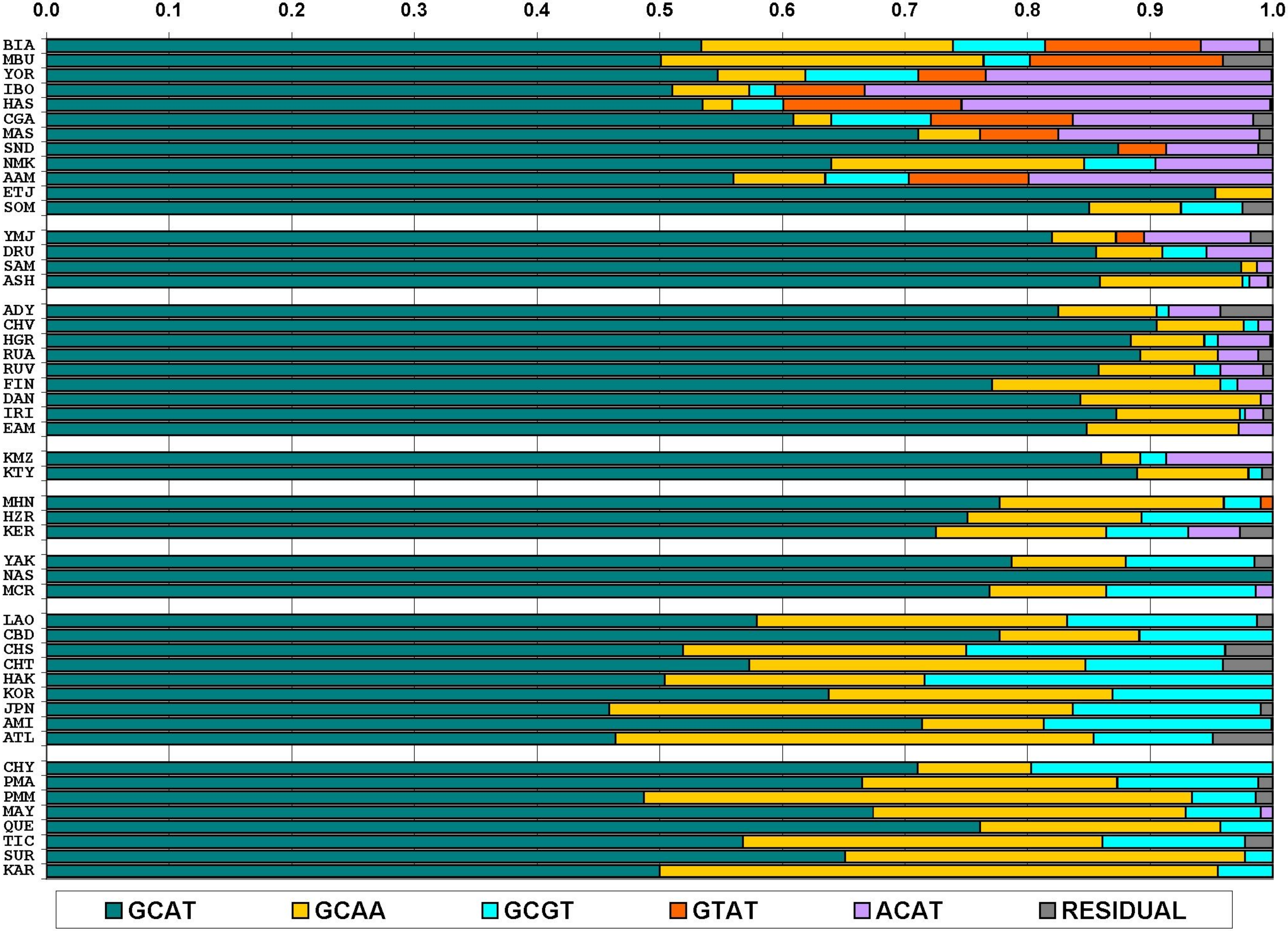
Supplemental Figure S3: Haplotype frequencies for Core-A upstream markers (1-5) in 50 population samples.

The 5 common haplotypes with frequencies of 5% or more are shown. Rare haplotypes are combined in residual class.



Supplemental Figure S4: haplotype frequencies for Core B (markers 6 through 9).

The 4 common haplotypes with frequencies of 5% or more are shown. Rare haplotypes are combined in residual.



Supplemental Figure S5: Haplotype frequencies for Core-C (markers 10-11)

Three of 4 haplotypes occur at common frequencies of 5% or more.