Full Title:

Contrasting exome constancy and regulatory region variation in the gene encoding CYP3A4: an examination of the extent and potential implications

Short Title

CYP3A4 variation: extent and potential implications.

Authors

Olivia J. Creemer[a],[a], Rosemary Ekong[a], Ayele Tarekegn[a,b], Christopher Plaster[a], Yuval Itan[a], Endashaw Bekele[a], Neil Bradman[c]

[a] The Centre for Genetic Anthropology, Research Department of Genetics, Evolution and Environment, University College London, London, UK, [b] Addis Ababa University, Addis Ababa, Ethiopia. [c] The Henry Stewart Group, 28-30 Little Russell Street, London WC1A 2HN

Correspondence to Dr Olivia J. Creemer, The Centre for Genetic Anthropology, Research Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, UK. Email: Oliviacreemer@gmail.com.

Authors thank all DNA sample donors. This study was funded in part by a charitable trust and a company of which Neil Bradman is a trustee and director and shareholder respectively. Neither tthe trust nor the company has any intellectual property or other rights with respect to the research. All authors have no conflict of interest to declare.

Abstract

Background and objectives: CYP3A4 expression varies up to 100-fold among individuals, and, to date, genetic causes remain elusive. As a major drug metabolising enzyme, elucidation of such genetic causes would increase the potential for introducing personalised dose-adjustment of therapies involving CYP3A4 drug substrates. The fetal CYP3A isoform, CYP3A7, is reported to be expressed in approximately 10% of European adults and may contribute to the metabolism of endogenous substances and CYP3A drug substrates, yet little is known about the distribution of variants associated with adult expression.

Methods: We resequenced the exons, flanking intronic regions, regulatory elements and 3'UTR of *CYP3A4* in five Ethiopian populations and incorporated data from The 1000 Genomes Project. Using bioinformatic analysis we assessed likely consequences of observed *CYP3A4* genomic variation. We also undertook the first extensive geographic survey of alleles associated with adult expression of CYP3A7: *CYP3A7*1C* and *CYP3A7*1B*.

Results and conclusions: Ethiopia contained 60 *CYP3A4* variants (26 novel) and more variants >1% than all non African populations combined. No non synonymous mutation was found in homozygous form or >2.8% in any population. 79% of haplotypes contained 3' UTR and/or regulatory region variation with striking pairwise population differentiation highlighting potential for inter-ethnic variation in CYP3A4 expression. Conversely, coding region variation revealed that there is unlikely to be great inter-ethnic variation in the type of CYP3A4 protein produced. *CYP3A7*1C,* was found at up to 17.5% in North African populations and in significant LD with *CYP3A5*3* indicating that adult expression of the fetal isoform is likely to be accompanied by reduced or null expression of CYP3A5.

Key words

Cytochrome P450 3A, CYP3A4, CYP3A5, CYP3A7, cytochrome P450 3A, drug metabolism, Ethiopia, The 1000 Genomes Project, pharmacogenetics, hormone-sensitive cancer.

Introduction

The human CYP3A subfamily is one of the most versatile of the biotransformation systems. Expressed predominantly within the liver and intestine, CYP3A enzymes contribute to the first pass systemic metabolism of more drugs than any other P450, an estimated 60% of those available [1,2]. Their associated genes are located in a cassette at chromosome 7q21-22. CYP3A enzymes also facilitate the metabolism and biosynthesis of other xenobiotics and endogenous substances including cholesterols, bile acids, vitamin D and steroid hormones [2, 3].

There are four functional CYP3A isoforms in humans: CYP3A4, CYP3A5, CYP3A7 and CYP3A43 [4, 5]. CYP3A4, the major adult isoform, comprises the bulk of adult hepatic P450 and up to 95% of hepatic CYP3A content [2, 6-8].

The predominant CYP3A in the foetal liver, CYP3A7 is down regulated after birth and gradually replaced by CYP3A4, contributing on average just 2% to the adult hepatic *CYP3A* mRNA pool [7]. However, recently it has been accepted that CYP3A7 expression in the adult liver and intestine is polymorphic; with 11% of adult Europeans reported to belong to a distinct subgroup with a high CYP3A7 expression phenotype [7, 9-11]. This has been attributed to two regulatory region polymorphisms; *CYP3A7*1B* and *CYP3A7*1C* [10]. In the foetal liver CYP3A7 is a major source of 16α-hydroxydehydroepiandrosterone (16α-OH DHEA) and its sulphated analog 16α-OH DHEAS, estriol precursors fundamental during pregnancy [12, 13]. CYP3A7 also has high catalytic activity for 16α-hydroxylation of estrone sulphate, the biologically inactive sulphated form of estrone [14, 15].

CYP3A5 is polymorphic in the adult due to non functional alleles, namely *CYP3A5*3*, *CYP3A5*6* and *CYP3A5*7*, that vary in frequency among ethnic groups. The most common, *CYP3A5*3*, is found at highest frequency in non-African populations such that only approximately 20% of Europeans exhibit hepatic expression [1, 8, 16]. CYP3A5 contributes approximately 4% to hepatic CYP3A content in *CYP3A5*3* homozygotes [16-18].

CYP3A43, a fourth member of the sub-family is a minor isoform that undergoes extensive alternative splicing such that most transcripts are non functional [5, 19, 20].

Hepatic and intestinal CYP3A expression may vary up to 100-fold [1, 21-23]. 60-90% of this variation is hypothesised to be due to genetic variation within the CYP3A locus [22] and may have considerable impact on drug pharmacokinetics, altering drug safety and efficacy particularly for substrates metabolised primarily by CYP3A enzymes and with narrow therapeutic indices

[24]. Identification of CYP3A variants causative of clinically variable phenotypes could offer potential for improving predictions of drug safety and efficacy [1, 24, 25].

As the major adult CYP3A isoform, variable expression of CYP3A4 is of relevance in healthcare. To date, genetic causes of wide variation in CYP3A4 expression remain elusive and their elucidation would separate genetic causes of variation from environmental determinants and increase the potential for introducing personalised dose-adjustment of therapies involving CYP3A4 drug substrates.

Variable expression of CYP3A enzymes has also been implicated in the risk of hormone sensitive cancers of the breast and prostate. CYP3A enzymes have numerous roles in steroidogenesis; CYP3A4, CYP3A5 and CYP3A43 enzymes catalyse the 2β-, 6β -, and 15β - hydroxylation of testosterone, leading to the formation of less biologically active metabolites [26, 27]. *CYP3A4* and *CYP3A5* gene variants have been associated with higher-grade prostate cancer tumours in ‘Caucasians’ and African Americans [28-31], and the early onset of puberty, a known risk factor for breast cancer [32].

Significantly reduced levels of steroid hormones associated with *CYP3A7*1C* has prompted further speculation as to whether adult expression of CYP3A7 is also associated with risk of such hormone sensitive cancers of the breast and prostate [33]. Unfortunately, little is known about the distribution of either *CYP3A7*1B* and *CYP3A7*1C*. Reported frequencies are predominantly from studies of European sample sets [10, 11, 16] and data available from The International HapMap project and The National Center for Biotechnology Information are currently incomplete or reported as ‘suspect’ [34, 35].

Of interest, given the extremely wide range of inter-individual expression levels, paucity of proposed environmental determinants and its importance as a drug metabolising enzyme, there are only two reports of *CYP3A4* null alleles: *CYP3A4*20*, found in four heterozygous individuals from the same family [36] and *CYP3A4*6* found in two heterozygous individuals [37, 38]. Thus, unlike CYP3A5, for which many non-expressing individuals are known, CYP3A4 may be essential for maintaining life. Excluding this study, 81 allelic variants have been reported in the genomic region extending from the most distal upstream enhancer to the 3’ UTR. Only two non-synonymous (NS) variants have been identified in homozygous form [38, 39].

Establishing the nature and extent of variation in *CYP3A4* and the implications for healthcare is important. We contribute to this undertaking by: a) re-sequencing the gene’s exons, flanking intronic regions, regulatory elements and 3’UTR in populations likely to evidence substantial genomic variation and b) bioinformatic analysis of the frequencies, distribution and likely consequences of observed genomic variation. Our findings should provide a firm foundation for undertaking future *in vivo* and *in vitro* expression analysis. We adopt the approach of Browning et al, (2010) [40] by characterising *CYP3A4* genomic variation within five diverse ethnic groups from Ethiopia and extend it to include data on 14 other populations from around the world using sequencing data obtained from The 1000 Genomes Project. In addition, because it has been reported that *CYP3A7* alleles are expressed in up to 11% of European adults we undertook the first extensive geographic survey of alleles associated with adult expression of this gene (*CYP3A7*1C* and *CYP3A7*1B*).

Finally, since CYP3A enzymes overlap considerably in their metabolic activities involving both endogenous and exogenous substances and have a) multiple roles in steroidogenesis (with the consequence that different combinations of *CYP3A* alleles can produce a range of phenotypes and b) potential implications for tacrolimus and cyclosporine dose in post-transplant care we analysed possible interactions of common polymorphisms within the *CYP3A7*, *CYP3A5* and *CYP3A4* cassette.

The association between the *CYP3A5*3* genotype and tacrolimus dose is well established [41-45] with *CYP3A5*1A/*1A* (expresser) and *CYP3A5*1A/*3* individuals exhibiting clearance rates for tacrolimus 25-45% greater than *CYP3A5*3/*3* (non-expresser) individuals, thus requiring significantly higher drug doses to achieve target drug concentration and prevent graft rejection [reviewed in 46].

Methods

Samples

Ethiopian DNA samples were collected and prepared according to [40]. Samples comprised Afar (n = 76), Anuak (n =76), Maale (n = 76), Oromo (n = 76) and Amhara (n = 77).

CYP3A4 genotypes of 1094 individuals (61 African Americans, 89 British, 60 Colombian, 87 Europeans, 93 Finnish, 97 Han Chinese, 89 Japanese, 97 Luhya from Kenya, 66 Mexican Americans, 55 Puerto Ricans, 100 Southern Han Chinese, 14 Spanish, 98 Tuscans, 88 Yoruba) from May 2011 sequencing calls (sequence and alignment release 20101123) of The 1000 Genomes Project were analysed. This release is based on the GRCh37 assembly of the human genome and is available in VCF4.0 format (http://www.1000genomes.org/data).

For the geographical survey of *CYP3A7*1B* and *CYP3A7*1C*, samples from the collection maintained by The Centre for Genetic Anthropology collected as described for the Ethiopian datasets were analysed. They comprised: 131 Algerians, 65 Anatolian Turks, 71 Armenians, 83 Ashkenazi Jews, 81 British, 287 Cameroonians, 51 Congolese, 83 Friesians, 132 Ghanaians, 244

Malawians, 80 Moroccan Berbers, 88 Moroccan Jews, 94 Mozambiquans, 79 Nigerians, 123 Senegalese, 62 Sephardic Jews, 37 South African Bantu speakers, 229 Sudanese, 48 Tanzanians, 39 Ugandan Bantu speakers, 113 Yemeni and 41 Zimbabweans. Details of collection locations are provided in Supplementary Material S1, Table S1.

Individuals carrying alleles associated with adult expression of CYP3A7 were further genotyped to determine their predicted CYP3A5 expression status. *CYP3A4/CYP3A7* haplotypes were inferred by pooling Ethiopian *CYP3A7* and *CYP3A4* regulatory region variation data.

Amplification and sequencing of *CYP3A4*

Amplification and sequencing conditions for all exons, flanking regions of adjacent introns, the 3' UTR and 5' regulatory regions are described in Supplementary Material S2. All 13 exons of *CYP3A4*, flanking regions of adjacent introns and the 3' UTR were sequenced. Upstream of the gene, approximately 500 nucleotides of the promoter were sequenced in addition to the two upstream enhancers: the xenobiotic responsive enhancer module (XREM) at -7.7 kb to -7.9 kb upstream [47], and the constitutive liver enhancer module 4 (CLEM4) at -10.9 kb to -11.4 kb upstream [48]. All primers are detailed in Supplementary Material S2, Table S2.

Amplification of *CYP3A7*

*CYP3A7*1C* comprises seven SNPs that occur together in complete Linkage Disequilibrium (LD) due to a gene conversion event: -232A>C (rs45446698), -262T>A (rs11568826), -270T>G (rs11568825), -281T>C (rs45467892), -282T>C (rs45575938), -284T>A (rs4594802) and -291G>T (rs11568824) [16]. The seven SNPs are hereafter referred to collectively as *CYP3A7*1C*. *CYP3A7*1B* (rs45465393) is a C>T transition 314 nucleotides upstream of the translation start site [16].

PCR amplification was undertaken as described in Supplemental Material S2, using the primers 5'–GGCATAGGTAAAGATCTGTAGGCAT-3' and 5'–AACTTTGGGATGCTGAGCAG-3' at an annealing temperature of 56°C generating an amplicon of 692 nucleotides. 492 nucleotides of this amplicon were sequenced as described in Supplemental Material S2 using the primers 5'–GGCATAGGTAAAGATCTGTAGGCAT-3' and 5'-ATCTCATCCCAAACTTGGCCG-3'.

Predicting CYP3A5 expression status

Genotyping for *CYP3A5*3* (rs776746), *CYP3A5*6* (rs10264272) and *CYP3A5*7* (rs41303343) within Moroccan Berbers and Algerians was undertaken by Kbioscience UK (www.kbioscience.co.uk). 12 µl of genomic DNA (per sample) at a concentration of 3.3 ng/µl was provided for genotyping in 96-well plates.

Statistical Analysis

For *CYP3A4* data, pairwise LD was assessed by the D' parameter and the logarithm (base 10) of odds score (LOD) using Haploview Software v4.2 [49]. *CYP3A4* rare variants (minor allele frequency <0.001) were excluded. LD was assessed using the pooled world dataset, pooled Ethiopian populations (Ethiopian dataset) and each individual population separately. For *CYP3A7* and *CYP3A5* data, LD was measured using Arlequin 3.11 Software [50]. Arlequin 3.11 Software was also used to test for departure of genotype frequencies from Hardy Weinberg Equilibrium (HWE), calculate hierarchical F_{ST} statistics and infer genetic distances between populations as represented by population pairwise F_{ST} values. The following were calculated using DNAsp 5.10 software [51]: Tajima's D [52] and Fu and Li's [53] tests of neutrality, gene diversity (h), nucleotide diversity (π), and Fisher's Exact Test. Analysis of variance (ANOVA) and Tukey's multiple comparison tests were undertaken using the Minitab 15 Software (www.minitab.com/en-US/default.aspx) and Sign Tests undertaken using Graphpad (www.Graphpad.com).

Haplotypes were inferred from unphased population genotype data using Phase Software version 2.1 [54]. Principal Coordinate Analysis was performed using the R statistical package [55] on pairwise similarity matrices, where similarity is quantified as being equal to the value of genetic distance subtracted from 1 (1-F_{ST}).

Inferring haplotypes of *CYP3A7*1B*, *CYP3A7*1C* and *CYP3A4*

Ethiopian data for *CYP3A7* and the regulatory region of *CYP3A4* were combined and haplotypes inferred as described above. Only samples with complete genotype data for all variants were included. In total 295 samples with complete *CYP3A7* genotype data and twelve of the fourteen *CYP3A4* regulatory region variants were analysed.

Bioinformatic analysis

Putative effects of amino acid substitutions were determined using PolyPhen 2 Software [56]. Human Splicing Finder (HSF) software (http://www.umd.be/HSF/) [57] was used to predict the putative impact of variants within 20 nucleotides of an exon/intron boundary.

CYP3A4 5' regulatory region variants and *CYP3A7*1D* (-92G>A) were analysed *in silico* to predict the putative impact of mutations on transcription factor binding (TFB) sites using MATCH™ version 8.3, provided by BIOBASE (http://www.biobase-international.com/) and the SNPinspector tool within Matinspector provided by Genomatix. 60 nucleotides flanking each allele of a mutation were used as the sequence input with vertebrate matrices selected and a minimum MSS score of 0.8 used as the cut off for matches. SNPinspector is only able to analyse single nucleotide changes thus insertions and deletions of more than one nucleotide were only analysed using MATCH.

Results

CYP3A4 variation within Ethiopia and populations of The 1000 Genomes Project

60 *CYP3A4* polymorphic sites were found in Ethiopia (Table 1), including 26 novel *CYP3A4* variants. 89 polymorphic sites were reported by The 1000 Genomes Project (Table 2).

Ethiopian populations (752 chromosomes) contained twice as many variants at frequencies >1% than all non-African populations combined (1122 chromosomes) and fourteen novel *CYP3A4* variants (excluding singletons). Eight variants (excluding singletons) were private to a single Ethiopian population. The Luhya and Finnish were the only other populations found to contain private alleles (three and two each respectively). 26164T>C was the only variant that deviated significantly from HWE (within the Maale, χ^2 17.74, P <0.001 after Bonferroni correction).

Twenty four coding region variants were present in the combined Ethiopian and 1000 Genomes dataset (referred to hereon as the world dataset), of which 15 were found on more than one chromosome. No NS mutation was found in homozygous form or at a frequency >2.8% in any population. All residues reported as part of the CYP3A4 active site, interacting with the haem group or implicated in binding, stereo specificity and cooperativity of the enzyme were monomorphic. One residue of the CYP3A4 phenylalanine cluster was altered by a singleton variant in Mexican Americans (15713T>C). A novel variant, 16883T, found on a single chromosome in Ethiopia was predicted as creating a premature stop codon within exon 8.

Within Ethiopia, the majority of *CYP3A4* loci had pair wise D' values of 1 indicating complete LD (Fig.1) (LD plots for all other populations can be found within Supplemental Material S4). The highest level of LD was observed toward the 3' end of the gene, occurring between the 3' UTR and exon 7. A line of LD was observed across *CYP3A4* involving pairs of loci that included -392A>G (*CYP3A4*1B*).

Bioinformatic analysis of *CYP3A4* variation

Excluding singletons, PolyPhen 2 predicted seven NS variants to cause benign changes to protein structure/function (Table 3).

Haplotypes inferred using only coding region variation, so as to restrict the haplotype set to those most likely to affect protein structure/function (excluding singletons), revealed sixteen coding region haplotypes (Fig. 2). Based on these haplotypes, 88% of *CYP3A4* chromosomes contained no coding region variation. Haplotypes predicted as damaging to CYP3A4 structure/function never exceeded 2.2% in any population (Table 4)

Three intronic variants were predicted to impact *CYP3A4* splicing: 14319A, 22050T and 20239A (Table 5). 20239A was particularly interesting due to its presence within all populations and high frequency (40-89%) in those of recent African origin. 20239A was in high LD with -10502C>T, *CYP3A4*1B*, *CYP3A4*1D* and -10852C>T indicating that any functional impact of this variant would likely be observed in combination with that of regulatory region variation. 14319A and 22050T did not exceed a frequency of 2% in any population.

Five regulatory region variants were hypothesised to impact *CYP3A4* transcriptional regulation and were present at polymorphic frequencies in Ethiopia (Table 6). Each was predicted to destroy/create functional/putative TFB sites upstream of the gene with potential implications for transcription factor (TF) binding. Haplotypes containing these variants were representative of almost half (43%) of the Ethiopian chromosomes sequenced (Table 7).

CYP3A4 haplotypes

203 *CYP3A4* haplotypes were inferred from the world dataset (excluding singletons), Supplementary Material S3, Fig. S1 and Table S3. Over half (103) of *CYP3A4* haplotypes were exclusive to Ethiopia. 99% of haplotypes contained multiple variants. 182 haplotypes contained variants that were respectively: within the regulatory region (n=134), within the 3' UTR (n=84), a putative splice variant (n=142), or were NS and predicted to impact protein structure/function (n=4). Thus, 90% of haplotypes contained variation with potential to impact expression and/or protein structure/function. Particularly interesting was the presence of regulatory region variation and/or the putative splice variant, 20239A, on >60% of haplotypes in comparison to coding region variation which was found on just 9% of haplotypes. This indicates that altered transcription and alternative splicing may be more important in understanding variable CYP3A4 expression than NS mutations.

Twenty one haplotypes were defined by intronic variation not known or predicted to impact CYP3A4 expression/function and can thus be assumed as similar in expression and function to *CYP3A4*1A* until shown otherwise. These haplotypes were representative of 87% of European, 79% of Asian, 61% of Latin American, 26% of African and 22% of African American chromosomes indicating that European populations are most likely to express a CYP3A4 protein similar in function to *CYP3A4*1A*. This is inclusive of haplotype 94, (defined by 19762T, intron 7), the most frequent *CYP3A4* haplotype found.

Haplotype 94 ranged in frequency from 5% in the Luhya to 86% in the European CEU, > 80% of European chromosomes were established as haplotype 94 compared to a maximum of 35% for African chromosomes. The Luhya, Yoruba and Anuak were the only populations within which haplotype 94 was not the most frequent haplotype. Haplotypes most frequent within these populations (2, 16, 49 and 150) all contained the promoter variant -392G (*CYP3A4*1B*) and the putative splice variant 20239A

Intragenic and intra population diversity

For the combined world dataset, mean gene (h) and nucleotide (π) diversity were consistently highest for *CYP3A4* introns (h 0.59 ± 0.26, π 0.00053 ± 0.00030), lowest for exons (h 0.08 ± 0.07, π 0.00006 ± 0.00005) and significantly different among the different gene regions (ANOVA: $P<0.001$) (Fig.3). *CYP3A4* introns were significantly more diverse than all other gene regions for both π and h, and *CYP3A4* exons significantly less diverse than the 5' regulatory region for h (Tukey's post-hoc analysis $P<0.05$). Fig. 4 depicts *CYP3A4* h and π for individual populations, the world dataset and the combined Ethiopian dataset. Consistent with the hypothesis that modern humans originated within Africa, populations of recent African origin were noticeably more diverse than non-African populations. This was true for all regions of *CYP3A4*. The high diversity of *CYP3A4* exons within populations of recent African origin could indicate that exonic variation is better tolerated within Africa, or may simply be a consequence of these older populations having had more time within which to accumulate coding region variation.

The diversity of Ethiopian populations was striking; populations placed in order of ascending π revealed Ethiopians to be the five most diverse populations for the 3' UTR, and that Ethiopians were the majority among the top five most diverse populations for all other gene regions, Supplementary Material S3, Table S4.

How different are populations based on *CYP3A4* haplotypes?

The majority of populations were significantly differentiated when whole gene haplotypes were considered ($P<0.0001$), Supplementary Material S3, Figure S2; Ethiopian populations were significantly differentiated from all populations of The 1000 Genomes Project and there was considerable pairwise differentiation between populations of recent African origin and those of Europe (Fig. 5).

Analysing each individual gene region separately there was a notable difference in pairwise population differentiation based on *CYP3A4* coding region variation compared to that for the 3’ UTR and 5’ regulatory region , Fig. 5, see also Supplementary Material S3, Figures S2-S5. Low population differentiation based on coding region variation indicates that unless a consequence of regulatory region or alternative splicing substantial ethnic variation in the type of CYP3A4 protein produced is unlikely. In contrast, because variation within the 3’ UTR and 5’ regulatory region has the potential to impact CYP3A4 expression, observed differences among populations suggest potential for ethnic variation in the amount of CYP3A4 expressed.

Testing for selection

Tajima’s D and Fu and Li’s D* and F* were negative for all populations (Supplementary Material S3, Table S5), a Sign test on the individual results yielded $P<0.0001$ for all three statistics.

Purifying selection was the prior hypothesis given the rarity of homozygous *CYP3A4* NS mutations and the lack of reported homozygous null alleles. Resequencing *CYP3A4* did not reveal an excess of *CYP3A4* NS mutations, rather it confirmed that *CYP3A4* exons are significantly more conserved than non coding regions of the gene. It is consequently unlikely that positive selection underlies the observed negative deviations rather than purifying selection.

Distribution of variants associated with adult expression of CYP3A7

Variation found upstream of *CYP3A7*

*CYP3A7*1C* was found in African, Middle Eastern and European populations ranging in frequency from <1% to a maximum of 17.5% in Moroccan Berbers. *CYP3A7*1B* was observed as a singleton within Friesian and British populations confirming the previously reported low frequency distribution of this variant (Table 8).

Eight additional variants were found upstream of *CYP3A7*, six were singletons (not shown) and one was the *CYP3A7*1D* (-49G>A) variant previously reported at 1% in Caucasians [16]. *CYP3A7*1D* was the most frequent *CYP3A7* variant found, ranging in frequency from 1% in Moroccan Jews to 22% in Cameroonians of Mayo Darle and was the only variant found in all African populations studied (Table 8).

Significant deviation from HWE was observed when populations were considered separately. *CYP3A7*1C* deviated significantly within the Senegalese Wolof (χ^2 63) and Armenians (χ^2 16.7), and *CYP3A7*1D* within the Mayo Darle

(χ^2 212.04) (Bonferroni correction for multiple tests, *P*<0.001). No variant was identified within all populations.

Are predicted adult expressers of CYP3A7 also predicted to express CYP3A5?

Algerian and Moroccan Berber individuals carrying one or more *CYP3A7*1C* alleles were genotyped for *CYP3A5*3*, *CYP3A5*6* and *CYP3A5*7* (Table 9). *CYP3A5*3* was the most common *CYP3A5* null allele at 86% in both populations. Interestingly, all individuals carrying one or more *CYP3A7*1C* alleles were found to carry at least one *CYP3A5*3* allele. More than half were predicted CYP3A5 non expressers (*CYP3A5*3/*3*) and one third were *CYP3A5*1A/*3* heterozygotes (predicted reduced CYP3A5 expression). Results thus provide good evidence that in these populations at least, adult expression of CYP3A7 is associated with reduced/null expression of CYP3A5.

Haplotype inference suggested *CYP3A7*1C* and *CYP3A5*3* are present on a shared haplotype background, haplotype 6, which is present at 7.1% and 17.9% within Algerians and Moroccan Berbers respectively (Table 10). Significant LD was observed between the two alleles (D' 0.85, χ^2 6.15, *P*<0.02, 1 df) with evidence of low frequency (<1%) recombination between *CYP3A5*3* and *CYP3A7*1C*, as determined by the four haplotype test.

Does *CYP3A7*1C* occur on the same haplotype background as *CYP3A4* regulatory region variants?

Combining Ethiopian *CYP3A7*1C* data with that for *CYP3A4* regulatory region variation revealed fifteen *CYP3A4/CYP3A7* haplotypes, three containing *CYP3A7*1C* (Table 11). *CYP3A7*1C* was found with *CYP3A4*1A* more frequently than any other *CYP3A4* regulatory region variant, occurring together on haplotype 13 at a total frequency of 3.1% in Ethiopia. Haplotypes 14 and 15 were found at <1% in Ethiopia. Haplotype 14, defined by *CYP3A7*1C* and -333T in significant LD (D' 1, χ^2 54.18, $P<0.001$, df 1) was exclusive to the Maale, at 1.6%. The functional impact of -333T remains to be determined but it was predicted here to destroy a putative glucocorticoid response element (GRE) in the *CYP3A4* promoter (Table 6). If functional, as has been hypothesised [58], the loss of this motif would likely reduce glucocorticoid receptor (GR) mediated induction of *CYP3A4*.

*CYP3A4*1B* was predicted to create a putative peroxisome proliferator response element (PPRE), a binding site for the peroxisome proliferator activated receptor (PPAR) family of nuclear receptors (Table 6). This is in agreement with reports that *CYP3A4*1B* increases expression of *CYP3A4* as PPARs are activating TFs [23, 59-61]. Adults carrying haplotype 15 could thus be expected to exhibit increased CYP3A4 and CYP3A7 expression (Table 11).

Potential implications of *CYP3A7*1D* for transcriptional activation of *CYP3A7*

Due to the widespread distribution and high frequency presence of *CYP3A7*1D*, *in silico* analysis was undertaken to identify any potential for this variant to alter *CYP3A7* TFB sites. Three putative TFB sites were identified in the region containing *CYP3A7*1D,* the putative impact of this variant on these sites is shown in Table 12.

In vitro work using foetal hepatocytes suggests potential for the direct regulation of *CYP3A7* by the glucocorticoid receptor (GR) [62]. If the GR is involved in *CYP3A7* regulation, then the functionality of the putative GRE created by *CYP3A7*1D* should be determined, the additional GRE may alter the *CYP3A7* expression profile in carriers of *CYP3A7*1D*.

AhR is a cytosolic TF [63] and activator of *CYP1A1*, but is not known to regulate any *CYP3A* [63, 65]. CCAAT enhancer binding proteins (C/EBPs) are regulators of *CYP3A* genes [65]. Therefore there is potential for the -51/-34 motif to be functional. This putative motif was not predicted to be disrupted by *CYP3A7*1D*.**Discussion**

Is CYP3A4 a protein essential for life?

Analysis of the world dataset lends weight to a hypothesis that homozygotes for *CYP3A4* null alleles either do not exist in the adult population or are extremely rare, an indication that *CYP3A4* expression may be essential. The datasets included in this study are collectively among the largest analysed for *CYP3A4* variation and the paucity of exonic variation is consistent with all current reports of *CYP3A4* variation as is the evidence for purifying selection.

Interestingly, unlike *CYP3A5* a near genomic neighbour with similar substrate affinities, no homozygotes for NS variants were observed and haplotypes predicted as damaging to CYP3A4 structure/function never exceeded 2.2% in any population. It is possible that polymorphic expression of CYP3A5 may have contributed to evolutionary conservation of *CYP3A4*, to compensate for reduced/null CYP3A5 enzyme activity. Alternatively, CYP3A4 expression may simply be exclusively responsible for the essential metabolism/biosynthesis of an unknown substrate. Indeed, unique endogenous roles do exist for CYP3A4, e.g. in Vitamin D biosynthesis [66, 67]. Determining why CYP3A4 might be an essential protein is beyond the scope of this study. However its results indicate possible directions for future research some of which are suggested below.

Implications for understanding variable expression of CYP3A4

Most previous studies have focused on identification and functional assessment of individual *CYP3A4* variants [e.g. 68-72]. Our results suggest that different combinations of alleles may have reinforcing or moderating effects making the haplotype rather than the SNP the preferred unit of interpretation.

99% of *CYP3A4* haplotypes identified were compound, often comprising combinations of variants with potential to impact CYP3A4 expression; 64% contained regulatory region variation, often in combination with the putative splice variant, 20239A (*CYP3A4*1G*). High LD between *CYP3A4*1G* and variants upstream of *CYP3A4* indicate that any functional impact of this polymorphism would likely be observed in combination with that of regulatory region variation. This may explain previous associations of *CYP3A4*1G* with both decreased [73-75] and increased CYP3A4 enzymatic activity [76]. NS variants including T185S and *CYP3A4*15* (previously suggested as causative of poor nifedipine metabolism [71]) were also found on compound haplotypes containing *CYP3A4*1G* and regulatory region variation. It is possible that the results of earlier *in vivo* functional studies, investigating individual gene variants, have been confounded by the impact of multiple variants present on the same haplotype background.

79% of haplotypes contained 3' UTR and/or regulatory region variation with striking pairwise population differentiation that highlights the potential importance of inter-ethnic variation in CYP3A4 expression, particularly between African and non-African populations. Conversely, much less differentiation of coding region variation was observed among groups suggesting that there is unlikely to be great inter-ethnic variation in the type of CYP3A4 protein produced.

Regulatory region variation may enhance or reduce gene expression. Five regulatory region variants hypothesised to impact *CYP3A4* transcriptional regulation were found at polymorphic frequencies within Ethiopian populations, *CYP3A4*1B*, -333C>T, -11185_-11186insTGT, -11131G>A and -10502T>C. Each was predicted to destroy/create putative/functional binding sites with hypothesised implications for TF binding and *CYP3A4* expression.

The impact of 3' UTR variation remains unknown. One study investigated MiRNA regulation of *CYP3A4* [77], reporting that MiRNA species are able to bind the *CYP3A4* 3' UTR with an impact on expression in human cell lines.

At present, ethnic differences in CYP3A4 expression are poorly characterised. Better understanding of the effects of polymorphisms in both the *CYP3A4* regulatory region and 3' UTR, if any, and their frequencies in different ethnic groups should contribute to better selection of subjects to clinical trials of CYP3A4 substrate drugs. Intergroup differences may be of particular importance in Sub Saharan Africa where genomic diversity is greatest.

Ethiopia is genetically diverse

Consistent with the hypothesis that modern humans migrated out of Africa via Ethiopia [78] much of *CYP3A4* variation reported in populations outside of Ethiopia was present within Ethiopian groups. Considerable novel *CYP3A4* variation was found with more than half of all *CYP3A4* entire gene haplotypes exclusive to Ethiopia. Measures of diversity consistently placed Ethiopian groups among the top five most diverse populations thus emphasising the considerable genetic diversity of Ethiopian ethnic groups.

Our work strongly favours Ethiopia as an attractive region in which to seek and evaluate novel variants. As a consequence, studies involving Ethiopian populations have the potential to make a substantial contribution to knowledge of genomic variation within clinically relevant genes.

Adult expression of CYP3A7

***CYP3A7*1C* and healthcare**

This study represents the most extensive ethno-geographic report of *CYP3A7* variants that have been associated with adult expression [11, 16, 22], implicated in the risk/progression of hormone sensitive cancers [33] and thought to have potential implications for the administration of CYP3A drug substrates.

*CYP3A7*1C* has been associated with significantly reduced levels of circulating steroid hormones that are known risk factors for and may play a role in the aetiology of hormone sensitive cancers [33]. *CYP3A7*1C* was at high frequency among populations of North Africa where, interestingly, average incidence rates (per 100,000 age standardised on world population) are 4.5 and 2-fold lower for prostate and breast cancer respectively (Algeria, Morocco, Libya, Tunisia and Egypt combined) when compared to Europe [79]. North African prostate cancer patients are also reported to have more favourable cancer characteristics than 'Caucasians' in general and Africans of Central Africa and the French West Indies [80].

The geographic distribution of *CYP3A7*1C* suggests that investigating the possible association between adult expression of CYP3A7 and the risk/progression of hormone sensitive cancer could yield interesting insight into the role CYP3A enzymes play in disease causation and progression. Ethnic background is recognised as playing a role in the detection, prevalence, pathologic state and prognosis of prostate cancer [81].

In the populations studied, predicted adult expression of CYP3A7 occurs on the background of a predicted reduced/null expression CYP3A5 phenotype. Adult expression of CYP3A7 may exceed CYP3A5 in individuals carrying both *CYP3A7*1C* and *CYP3A5*3* alleles [11, 82]. Thus, within the populations of this study where *CYP3A7*1C* is at high frequency, CYP3A7 expression may be quantitatively more important than that of CYP3A5. Consequently, association studies investigating the impact of *CYP3A7*1C* should be designed to take account of *CYP3A5* genotypes. CYP3A5 and CYP3A7 have different roles in steroidogenesis and different combinations of alleles of the two genes are likely to have different phenotype outcomes.

The distribution of *CYP3A7*1C* presented here may also be of practical relevance to the administration of CYP3A drug substrates. Adult expression of CYP3A7 is likely to be substantial in members of a number of ethnic groups. Thus, where the function of CYP3A7 is known to overlap with that of CYP3A4 and CYP3A5, *CYP3A7*1C* genotype should be taken into account when considering drug choice and dose. This is especially relevant to transplant pharmacology involving cyclosporine and tacrolimus. Genotyping for *CYP3A7*1C* may improve initial dosing strategies for transplant patients. This is of relevance for example in France where substantial numbers of the North African Diaspora live representing one of the largest ethnic minorities at approximately 9% of the general population (The World Factbook: France; www.cia.gov/library/publications/the-world-factbook/geos/fr.html).

Our results lend support to recent suggestions that *CYP3A* polymorphisms and their association with clinical conditions will be better understood by adopting a *CYP3A* haplotype approach [33] including multiple polymorphisms of *CYP3A4*, *CYP3A5*, and *CYP3A7*, particularly *CYP3A7*1C*, *CYP3A4*1B* and *CYP3A5*3* which have previously been associated with altered steroid hormone levels and cancer incidence [33].

CYP3A have multiple roles in steroidogenesis, thus there is likely to be difficulty in unravelling the functional impact of individual *CYP3A* polymorphisms without considering others present at the *CYP3A* locus. Data presented in this study suggest that a more rapid shift from researching gene variants to gene haplotypes is necessary if pharmacogenetic research is to properly address the genomic complexity that is becoming increasingly apparent. By doing so, pharmacogenetic research should better contribute to the development of personalised/stratified medicine and improved healthcare policy.

Finally, this study is consistent with that of [40] which reported extensive variation in *CYP1A2* within the peoples of Ethiopia and provides further support for the hypothesis that they represent a rich resource for assessing genomic variation, particularly when combined with data from worldwide populations as available in The 1000 Genomes project. Furthermore the presence of extensive variation outside the coding region, but not within it, supports hypotheses that: a) the protein CYP3A4 is essential for the maintenance of life and b) wide inter-individual differences in levels of expression are likely to be due in large measure to genomic variation in the promoter and, possibly, 3' UTR regions. It is possible to speculate that since, so far as we are aware, there are no reports of mortality in either adults or children associated with a *CYP3A4* mutation that if the protein is essential for life it plays an important, and as yet undetected, role either at or prior to conception or shortly thereafter notwithstanding that current opinion is that the enzyme is only expressed after birth.

References

1. Lamba JK, Lin YS, Schuetz EG, Thummel KE. Genetic contribution to variable human CYP3A-mediated metabolism. *Adv Drug Deliv Rev* 2002; **54**: 1271-1294.

2. Rendic S. Summary of information on human CYP enzymes: human P450 metabolism data. *Drug Metab Rev* 2002; **34:** 83-448.

3. Anzenbacher P, Anzenbacherova E. Cytochromes P450 and metabolism of xenobiotics. *Cell Mol Life Sci* 2001; **58:** 737-747.

4. Finta C, Zaphiropoulos PG. The human cytochrome P450 3A locus. Gene evolution by capture of downstream exons. *Gene* 2000; **260**: 13-23.

5. Finta C, Zaphiropoulos PG. Intergenic mRNA molecules resulting from trans-splicing. *Journal Biol Chem* 2002; **277:**5882-5890.

6. Shimada T, Yamazaki H, Mimura M, Inui Y, Guengerich FP. Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation

of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *Journal Pharmacol Exp Ther* 1994; **270**: 414-423.

7. Koch I, Weil R, Wolbold R, Brockmo J, Hustert E, Burk O, et al. Interindividual variability and tissue-specificity in the expression of cytochrome P450 3A mRNA. D*rug Metab Dispos* 2002; **32**:1108-1114.

8. Burk O, Wojnowski L. Cytochrome P450 3A and their regulation. *Naunyn Schmiedebergs Arch Pharmacol* 2004; **369**: 105-124

9. Schuetz JD, Beach DL, Guzelian PS. Selective expression of cytochrome P450 CYP3A mRNAs in embryonic and adult human liver. *Pharmacogenetics* 1994; **4**: 11-20.

10. Burk O, Tegude H, Koch I, Hustert E, Wolbold R, Glaeser H, et al. Molecular mechanisms of polymorphic CYP3A7 expression in adult liver and intestine. *J Biol Chem* 2002; **277** 24280-24288.

11. Sim SC, Edwards RJ, Boobis AR, Ingelman-Sundberg M. CYP3A7 protein expression is high in a fraction of adult human livers and partially associated with the CYP3A7*1C allele. *Pharmacogenet Genomics* 2005; **15**: 625-631.

12. Ohmori S, Nakasa H, Asanome K, Kurose Y, Ishii I, Hosokawa M, et al. Differential catalytic properties in metabolism of endogenous and exogenous substrates among CYP3A enzymes expressed in COS-7 cells. *Biochim Biophys Acta* 1998; **1380**:297-304.

13. Blackburn ST (2007). *Maternal, fetal & neonatal physiology. A clinical perspective*. 3rd Edition. United States of America. Saunders.

14. Lee AJ, Conney AH., Zhu BT. Human cytochrome P450 3A7 has a distinct high catalytic activity for the 16alpha-hydroxylation of estrone but not 17beta-estradiol. *Cancer Res* 2003; **63**: 6532-6536.

15. Miller KK, Cai J, Ripp SL, Pierce WM. Jr, Rushmore TH, Prough RA. Stereo- and regioselectivity account for the diversity of dehydroepiandrosterone (DHEA) metabolites produced by liver microsomal cytochromes P450. *Drug Metab Dispos* 2004; **32**:305-313.

16. Kuehl P, Zhang J, Lin Y, Lamba J, Assem M, Schuetz J, et al. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nat Genetics* 2001; **27**: 383-391.

17. Hustert E, Haberl M, Burk O, Wolbold R, He YQ, Klein K, et al. The genetic determinants of the CYP3A5 polymorphism. *Pharmacogenetics* 2001; **11**: 583-589.

18. Lin YS, Dowling AL, Quigley SD, Farin FM, Zhang J, Lamba J, et al. Co-regulation of CYP3A4 and CYP3A4 and contribution to hepatic and intestinal midazolam metabolism. *Mol Pharmacol* 2002; **62**:162-172.

19. Domanski TL, Finta C, Halpert JR, Zaphiropoulos PG. cDNA cloning and initial characterization of CYP3A4, a novel human cytochrome P450. *Mol Pharmacol* 2001; **59**: 386-392.

20. Gellner K, Eiselt R, Hustert E, Arnold H, Koch I, Haberl M, et al. Genomic organization of the human *CYP3A* locus: identification of a new, inducible CYP3A gene. *Pharmacogenetics* 2001; **11**: 111-121.

21. Westlind A, Lofberg L, Tindberg N, Andersson TB, Ingelman-Sundberg M. Interindividual differences in hepatic expression of CYP3A4: relationship to

genetic polymorphism in the 5'-upstream regulatory region. *Biochem Biophys Res Comm* 2001; **259**: 201-205.

22. Ozdemir V, Kalow W, Tang BK, Paterson AD, Walker SE, Endrenyi L. Evaluation of the genetic component of variability in CYP3A4 activity: a repeated drug administration method. *Pharmacogenetics* 2000; **10**:373-388.

23. Schirmer M, Toliat MR, Haberl M, Suk A, Kamdem LK, Klein K, et al. Genetic signature consistent with selection against the CYP3A4*1B allele in non-African populations. *Pharmacogenet Genomics* 2006; **16**; 59-71.

24. Wojnowski L, Kamdem LK. Clinical implications of CYP3A polymorphisms. *Expert Opin Drug Metab Toxicol* 2006; **2**:171-182.

25. Perera MA. The missing linage: what pharmacogenetic associations are left to find in CYP3A? *Expert Opin Drug Metab Toxicol* 2010; **6**:17-28.

26. Waxman DJ, Attisano C, Guengerich FP, Lapenson DP. Human liver microsomal steroid metabolism: identification of the major microsomal steroid hormone 6 beta-hydroxylase cytochrome P-450 enzyme. *Arch Biochem Biophys*1988; **263**:424–36.

27. Waxman DJ, Lapenson DP, Aoyama T, Gelboin HV, Gonzalez FJ, Korzekwa K. Steroid hormone hydroxylase specificities of eleven cDNA-expressed human cytochrome P450s. *Arch Biochem Biophys* 1991; **290**:160–6.

28. Rebbeck TR, Jaffe JM, Walker AH, Wein AJ, Malkowicz SB. Modification of clinical presentation of prostate tumors by a novel genetic variant in CYP3A4. J Natl Cancer Inst 1998; **90**:1225-1229.

29. Paris PL, Kupelian PA, Hall JM, Williams TL, Levin H, Klein EA, et al. Association between a CYP3A4 genetic variant and clinical presentation in African-American prostate cancer patients. *Cancer Epidemiol Biomarkers Prev* 1999; **8**: 901–905.

30. Tayeb MT, Clark C, Sharp L, Haites NE, Rooney PH, Murray GI, et al. CYP3A4 promoter variant is associated with prostate cancer risk in men with benign prostate hyperplasia. *Oncol Rep* 2002; **9**: 653-655.

31. Tayeb MT., Clark C, Haites NE., Sharp L, Murray GI, McLeod HL. CYP3A4 and VDR gene polymorphisms and the risk of prostate cancer in men with benign prostate hyperplasia. *Br J Cancer* 2003; **88**: 928-932.

32. Kadlubar FF, Berkowitz GS, Delongchamp RR, Wang C, Green BL, Tang G, et al. The CYP3A4*1B variant is related to the onset of puberty, a known risk factor for the development of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2003; **12**: 327-331.

33. Siemes C, Visser LE, de Jong FH, Coebergh JW, Uitterlinden AG, Hofman A, et al. Cytochrome P450 3A gene variation, steroid hormone serum levels and prostate cancer—The Rotterdam Study. *Steroids* 2010; **75**:1024-1032.

34. Musumeci L, Arthur JW, Cheung FS, Hoque A, Lippman S, Reichardt JK. Single nucleotide differences (SNDs) in the dbSNP database may lead to errors in genotyping and haplotyping studies. *Hum Mutat* 2010; **31**:67-73.

35. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, et al. Diversity of human copy number variation and multicopy genes. *Science* 2010; **330:641**-646.

36. Westlind-Johnsson A, Hermann R, Huennemeyer A, Hauns B, Lahu G, Nassr N, et al. Identification and characterization of CYP3A4*20, a novel rare CYP3A4 allele without functional activity. *Clin Pharmacol Ther* 2006; **79**: 339-349.

37. Hsieh K, Lin Y, Cheng C, Lai M, Lin M, Siest J, et al. Novel mutations of CYP3A4 in Chinese. *Drug Metab Dispos* 2001; **29**:268-273.

38. Fukushima-Uesaka H, Saito Y, Watanabe H, Shiseki K, Saeki M, Nakamura T, et al. Haplotypes of CYP3A4 and their close linkage with CYP3A5 haplotypes in a Japanese population. *Hum Mutat* 2004; **23**: 100.

39. Shchepotina EG, Vavilin VA, Goreva OB, Lyakhovich VV. Some mutations of exon-7 in cytochrome P450 gene 3A4 and their effect on 6beta-hydroxylation of cortisol. *Bull Exp Biol Medicine* 2006; **141**:701-703.

40. Browning SL, Tarekegn A, Bekele E, Bradman N, Thomas MG. CYP1A2 is more variable than previously thought: a genomic biography of the gene behind the human drug-metabolizing enzyme. *Pharmacogenet Genomics* 2010; **20**: 647-664.

41. Choi JH, Lee YJ, Jang SB, Lee JE, Kim KH, Park K. Influence of the CYP3A5 and MDR1 genetic polymorphisms on the pharmacokinetics of tacrolimus in healthy Korean subjects. *Br J Clin Pharmacol* 2007; **64**: 185-191.

42. Crettol S, Venetz JP, Fontana M, Aubert JD, Pascual M., Eap C.B. CYP3A7, CYP3A5, CYP3A4, and ABCB1 genetic polymorphisms, cyclosporine

concentration, and dose requirement in transplant recipients. *Ther Drug Monit* 2008; **30**: 689-699.

43. Fukudo M, Yano I, Yoshimura A, Masuda S, Uesugi M, Hosohata K, et al. Impact of MDR1 and CYP3A5 on the oral clearance of tacrolimus and tacrolimus-related renal dysfunction in adult living-donor liver transplant patients. *Pharmacogenet Genomics* 2008; **18**: 413-423.

44. Jun KR, Lee W, Jang MS, Chun S, Song GW, Park KT, et al. Tacrolimus concentrations in relation to CYP3A and ABCB1 polymorphisms among solid organ transplant recipients in Korea. *Transplantation* 2009; **87**:1225-1231.

45. Press RR, Ploeger BA, den Hartigh J, van der Straaten T, van Pelt J, Danhof M, et al. Explaining variability in tacrolimus pharmacokinetics to optimize early exposure in adult kidney transplant recipients. *Ther Drug Monit* 2009; **31**:187-197.

46. Barry A, Levine M. A systematic review of the effect of CYP3A5 genotype on the apparent oral clearance of tacrolimus in renal transplant recipients. *Ther Drug Monit* 2010; **32**:708-714.

47. Goodwin B, Hodgson E, Liddle C. The orphan human pregnane X receptor mediates the transcriptional activation of CYP3A4 by rifampicin through a distal enhancer module. *Mol Pharmacol* 1999; **56**:1329-1339.

48. Matsumara K, Saito T, Takahashi Y, Ozeki T, Kiyotani K, Fujieda M, et al. Identification of a novel polymorphic enhancer of the human CYP3A4 gene. *Mol Pharmacol* 2004; **65**: 326-334.

49. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**:263-265.

50. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 2007; **1**:47-50.

51. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009 **25**:1451-1452.

52. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; **123**:585-595.

53. Fu YX and Li WH. Statistical tests of neutrality of mutations. *Genetics* 1993; **133**: 693-709.

54. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; **68**: 978-989.

55. http://www.r-project.org/

56. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248-249.

57. Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; **37**: E67.

58. El-Sankary W, Bombail V, Gibson GG, Plant N. Glucocorticoid-mediated induction of CYP3A4 is decreased by disruption of a protein: DNA interaction distinct from the pregnane X receptor response element. *Drug Metab Dispos* 2002; **30**: 1029-1034.

59. Amirimani B, Walker AH, Weber BL, Rebbeck TR. RESPONSE: re: modification of clinical presentation of prostate tumors by a novel genetic variant in CYP3A4. *J Natl Cancer Inst* 1999; **91**: 1588-1590.

60. Ando Y, Tateishi T, Sekido Y, Yamamoto T, Satoh T, Hasegawa Y, et al. Re: Modification of clinical presentation of prostate tumors by a novel genetic variant in CYP3A4. *J Natl Cancer Inst* 1999; **91**: 1587-1590.

61. Amirimani B, Ning B, Deitz AC, Weber BL, Kadlubar FF, Rebbeck TR. Increased transcriptional activity of the CYP3A4*1B promoter variant. *Environ Mol Mutagen* 2003; **42**: 299-305.

62. Matsunaga T, Maruyama M, Harada E, Katsuyama Y, Sugihara N, Ise H, et al. Expression and induction of CYP3As in human fetal hepatocytes. *Biochem Biophys Res Commun* 2004; **318**: 428-434.

63. Ma C, Marlowe JL, Puga A. The aryl hydrocarbon receptor at the crossroads of multiple signaling pathways. *EXS* 2009; **99**: 231-257.

64. Androutsopoulos VP, Tsatsakis AM, Spandidos DA. Cytochrome P450 CYP1A1: wider roles in cancer progression and prevention. *BMC Cancer* 2009; **16**:187.

65. Rodríguez-Antona C, Bort R, Jover R, Tindberg N, Ingelman-Sundberg M, Gómez-Lechón MJ, et al. Transcriptional regulation of human CYP3A4 basal expression by CCAAT enhancer-binding protein alpha and hepatocyte nuclear factor-3 gamma. *Mol Pharmacol* 2003; **63**:1180-1189.

66. Gupta RP, Hollis BW, Patel SB, Patrick KS, Bell NH. CYP3A4 is a human microsomal vitamin D 25-hydroxylase. *J Bone Miner Res* 2004; **19**: 680-688.

67. Gupta RP, He YA, Patrick KS, Halpert JR, Bell NH. CYP3A4 is a vitamin D-24- and 25-hydroxylase: analysis of structure function by site-directed mutagenesis. J Clin Endocrinol Metab 2005; **90**:1210-1219.

68. Sata F, Sapone A, Elizondo G, Stocker P, Miller VP, Zheng W, et al. CYP3A4 allelic variants with amino acid substitutions in exons 7 and 12: evidence for an allelic variant with altered catalytic activity. *Clin Pharmacol Ther* 2000; **67**:48-56.

69. Dai D, Tang J, Rose R, Hodgson R, Bienstock RJ, Mohrenweiser HW, et al. Identification of variants of CYP3A4 and characterization of their abilities to

metabolize testosterone and chlorpyrifos. *J Pharmacol Exp Ther* 2001; **299**: 825-831.

70. Eiselt R, Domanski TL, Zibat A, Mueller R, Presecan-Siedel R, Hustert E, et al. Identification and functional characterization of eight CYP3A4 protein variants. *Pharmacogenetics* 2001; **11**: 447-458.

71. Lamba JK, Lin YS, Thummel K, Daly A, Watkins PB, Strom S, et al. Common allelic variants of cytochrome P4503A4 and their prevalence in different populations. *Pharmacogenetics* 2002; **12**: 121-132.

72. Murayama N, Nakamura T, Saeki M, Soyama A, Saito Y, Sai K, et al. CYP3A4 gene polymorphisms influence testosterone 6beta-hydroxylation. *Drug Metab Pharmacokinet* 2002; **17**:150-156.

73. Yuan R, Zhang X, Deng Q, Wu Y, Xiang G. Impact of CYP3A4*1G polymorphism on metabolism of fentanyl in Chinese patients undergoing lower abdominal surgery. *Clin Chim Acta* 2011; **412**: 755-760.

74. Dong ZL, Li H, Chen QX, Hu Y, Wu SJ, Tang LY, et al. Effect of CYP3A4*1G on the fentanyl consumption for intravenous patient-controlled

analgesia after total abdominal hysterectomy in Chinese Han population. *J Clin Pharm Ther* 2011; Early view: doi: 10.1111/j.1365-2710.2011.01268.x

75. Gao Y, Zhang LR, Fu Q. CYP3A4*1G polymorphism is associated with lipid-lowering efficacy of atorvastatin but not of simvastatin. *Eur J Clin Pharmacol* 2008; **64**: 877-882.

76. Hu YF, Tu JH, Tan ZR, Liu ZQ, Zhou G, He J, et al. Association of CYP3A4*18B polymorphisms with the pharmacokinetics of cyclosporine in healthy subjects. *Xenobiotica* 2007;**37**: 315-327.

77. Pan YZ, Gao W, Yu AM. MicroRNAs regulate CYP3A4 expression via direct and indirect targeting. *Drug Metab Dispos* 2009;**37**: 2112-2117.

78. Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, Gallego Romero I, et al. Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am J Hum Genet* 2012; **91**:83-96.

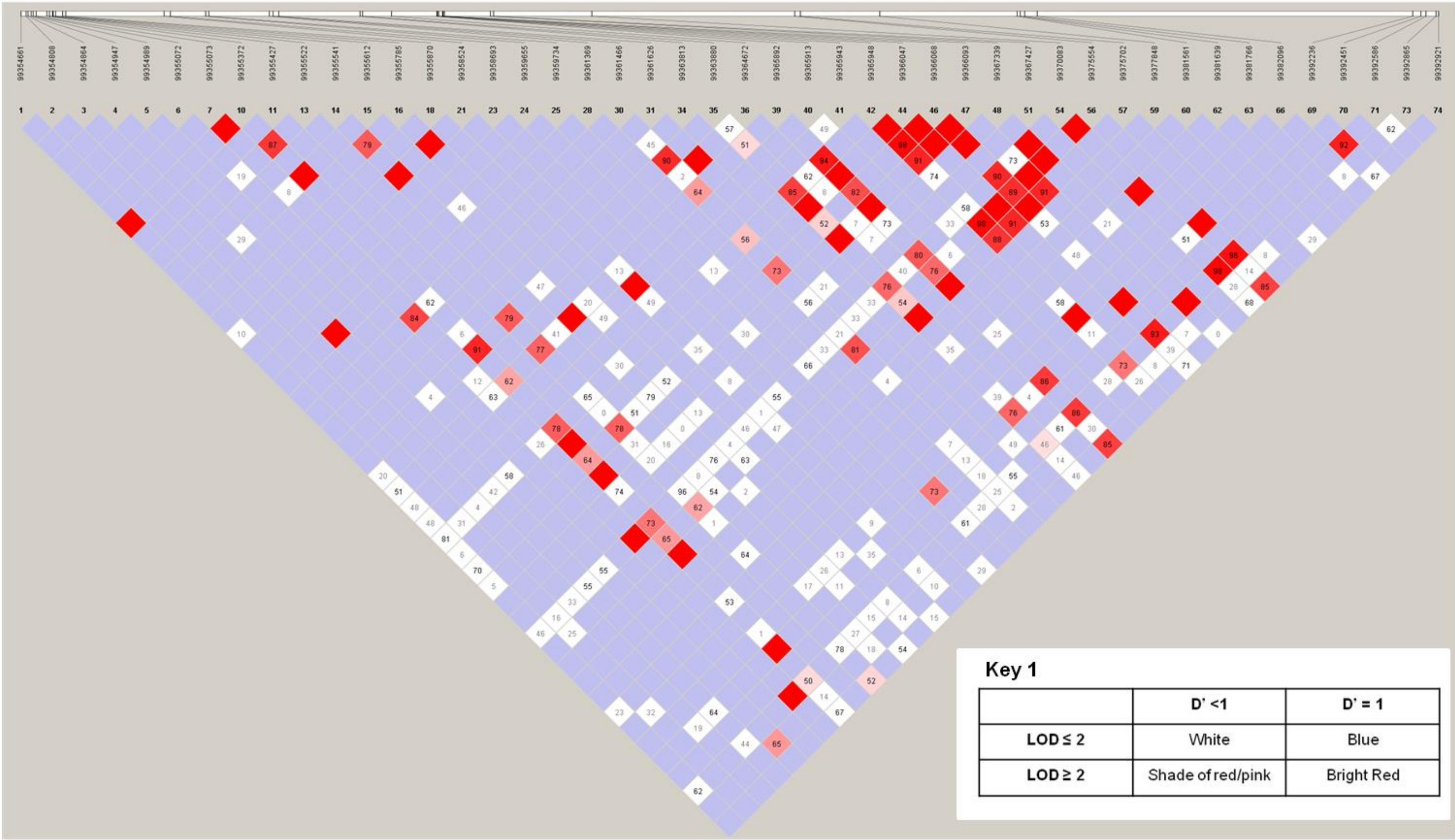
79. Zanetti R, Tazi MA, Rosso S. New data tells us more about cancer incidence in North Africa. *Eur J Cancer* 2010;**46**: 462-466.

80. Ravery V, Dominique S, Hupertan V, Ben Rhouma S, Toublanc M, Boccon-Gibod L. Prostate cancer characteristics in a multiracial community. *Eur Urol* 2008; **53**: 533-538.

81. Thompson EE, Kuttab-Boulos H, Witonsky D, Yang L, Roe BA, Di Rienzo A. CYP3A variation and the evolution of salt-sensitivity variants. *Am J Hum Genet* 2004; **75**:1059-1069.

82. Westlind-Johnsson A, Malmebo S, Johansson A, Otter C, Andersson TB, Johansson I, et al. Comparative analysis of CYP3A expression in human liver suggests only a minor role for CYP3A5 in drug metabolism. *Drug Metab Dispos* 2003;**31**: 755-761.

Fig.1 LD across *CYP3A4* in the Ethiopian dataset



The strength of LD is based on D' and LOD (logarithm of the odds) scores and increases with colour from white to blue to red as depicted in Key 1. At the top of the figure, the first row of numbers denotes the *CYP3A4* variant (Key 2), the second row indicates position of each variant on the chromosome. The space between loci is indicated by the white bar at the top of the plot. Numbers within squares are D' values (multiplied 100-fold). Empty squares indicate D' of 1.Singleton and low frequency (<0.0001) variants were excluded.

Key 2

Number	Chromosome pos.	*CYP3A4* variant	Location
1	99393120	-11386T>A	5' Upstream
2	99392921	-11185_-11186insTGT	5' Upstream
3	99392865	*-11131G>A*	5' Upstream
4	99392804	-11070G>C	5' Upstream
5	99392586	-10852C>T	5' Upstream
6	99392451	-10755_delG	5' Upstream
7	99392236	-10502C>T	5' Upstream
8	99389435	-7701A>G	5' Upstream
9	99382128	-394T>C	5' Upstream
10	99382096	-392A>G	5' Upstream
11	99381998	-333C>T	5' Upstream
12	99381919	-215T>A	5' Upstream
13	99381766	-62 C>A	5' Upstream
14	99381639	66 C>G	Exon 1
15	99381603	102 T>C	Intron 1
16	99381561	144 A>G	Intron 1
17	99377848	3857C>T	Intron 1
18	99377827	3873_3875delATT	Intron 1
19	99375702	6003G>A	Exon 3
20	99375554	6151G>A	Intron 3
21	99375547	6158 T>A	Intron 3
22	99370083	11522C>T	Intron 4
23	99367867	13838 G>C	Intron 4
24	99367727	13978 G>A	Intron 5
25	99367427	14278G>A	Exon 6
26	99367392	14313G>C	Exon 6
27	99367386	14319G>A	Intron 6
28	99367339	14366T>G	Intron 6
29	99366093	*15612C>G*	Exon 7
30	99366068	15637C>T	Exon 7
31	99366048	15658A>T	Exon 7
32	99366047	15658A>G	Exon 7
33	99365983	15722T>C	Exon 7
34	99365948	15757A>C	Intron 7
35	99365943	15762G>T	Intron 7
36	99365913	15792T>C	Intron 7
37	99365892	15813T>C	Intron 7
38	99365876	15829A>G	Intron 7

Number	Chromosome pos.	*CYP3A4* variant	Location
39	99365083	16622C>T	Intron 7
40	99364672	*17033C>T*	Intron 8
41	99363880	17824_17825delAT	Intron 9
42	99363813	17892C>G	Intron 9
43	99363760	17945C>T	Intron 9
44	99363731	17974A>G	Intron 9
45	99361626	20079T>C	Exon 10
46	99361466	20239G>A	Intron 10
47	99361387	20318G>C	Intron 10
48	99361369	20336T>C	Intron 10
49	99359911	21794A>C	Intron 10
50	99359800	21905C>T	Exon 11
51	99359734	21971A>G	Exon 11
52	99359655	22050C>T	Intron 11
53	99358693	23012G>A	Intron 11
54	99358615	23090C>T	Intron 11
55	99358524	23181T>C	Exon 12
56	99355975	25730A>G	Intron 12
57	99355957	25748C>T	Intron 12
58	99355870	25835G>A	Intron 12
59	99355789	25916T>G	Exon 13
60	99355785	25920T>A	Exon 13
61	99355612	26093T>C	3' UTR
62	99355541	26164T>C	3' UTR
63	99355522	26184C>T	3' UTR
64	99355490	26215C>A	3' UTR
65	99355427	26278C>T	3' UTR
66	99355372	26335T>C	3' UTR
67	99355278	26427C>T	3' UTR
68	99355185	26520C>T	3' UTR
69	99355073	26632C>T	3' UTR
70	99355072	26633G>A	3' UTR
71	99354989	26716_delT	3' UTR
72	99354947	26758T>C	3' UTR
73	99354864	26841C>A	3' UTR
74	99354808	26897A>T	3' UTR
75	99354661	27044C>G	3' UTR

Fig. 2 Haplotype inference across *CYP3A4* coding regions

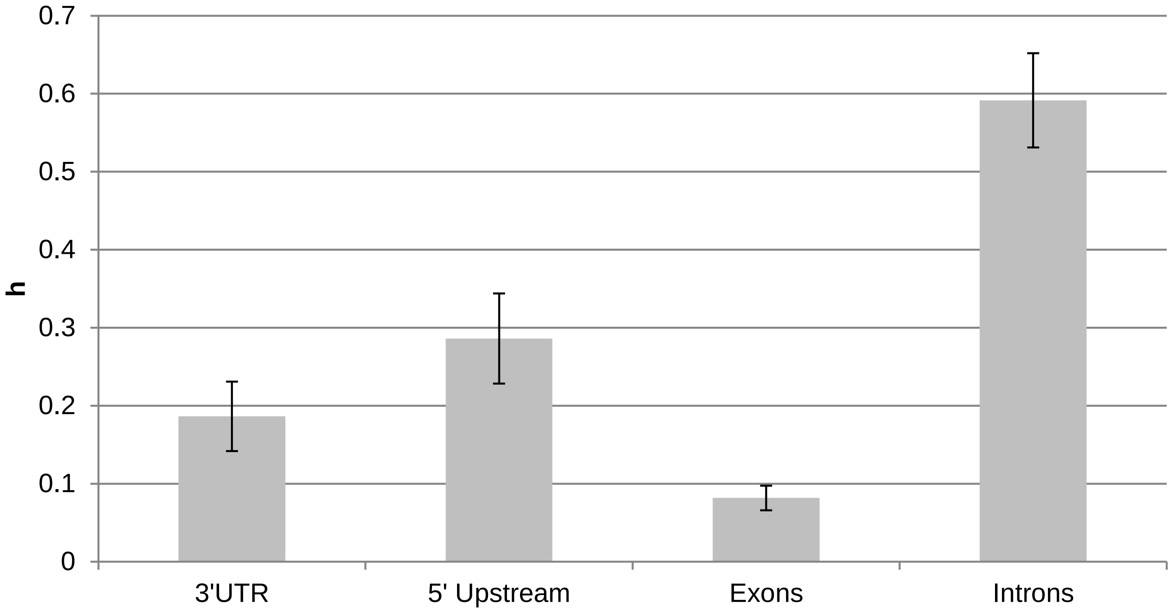
Nucleotide Change[1]	25920 T>A	25916 T>G	23181 T>C	21971 A>G	21905 C>T	20079 T>C	15722 T>C	15658 A>G	15658 A>T	15637 C>T	15612 C>G	14313 G>C	14278 G>A	6003 G>A	66 C>G	Predicted effect on CYP3A4 structure/function[3]
Exon	13	13	12	11	11	10	7	7	7	7	7	6	6	3	1	
Amino Acid Change	S495T		M445T	M395V	L373F	L293P	S222P		Q200H		T185S	D174H	R162Q	G56D		
Haplotype ID[2] 1																
2																Undamaged
3																Probably Damaged
4																Undamaged
5																Undamaged
6																Possibly Damaged
7																Undamaged
8																Undamaged
9																Possibly Damaged
10																Undamaged
11																Undamaged
12																Undamaged
13																Undamaged
14																Probably Damaged
15																Undamaged
16																Undamaged

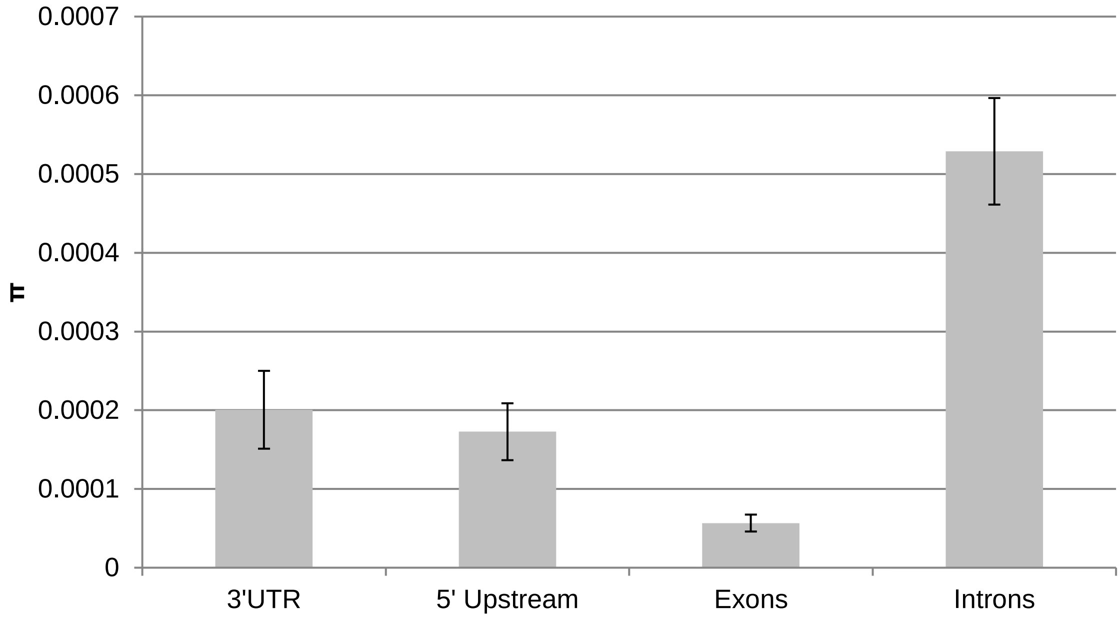
[1] Position from base A in initiation codon (A in ATG is considered as +1).

[2] White cell, allele observed in *CYP3A4*1A;* grey cell, derived allele.

[3] Predictions made using PolyPhen2 based on single amino acid alterations.

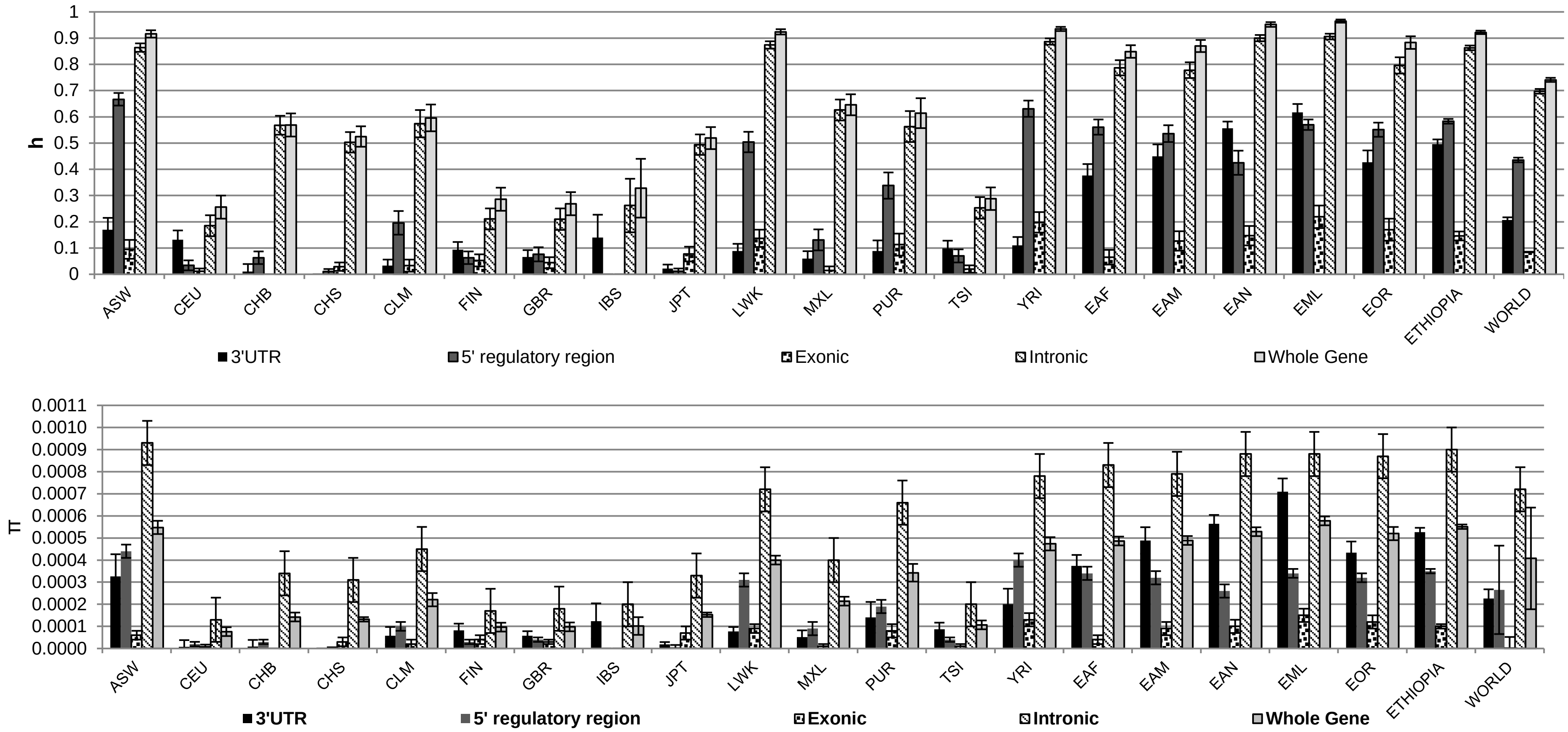
Fig. 3 Mean gene diversity (h) and nucleotide diversity (π) within regions of *CYP3A4* for the world dataset





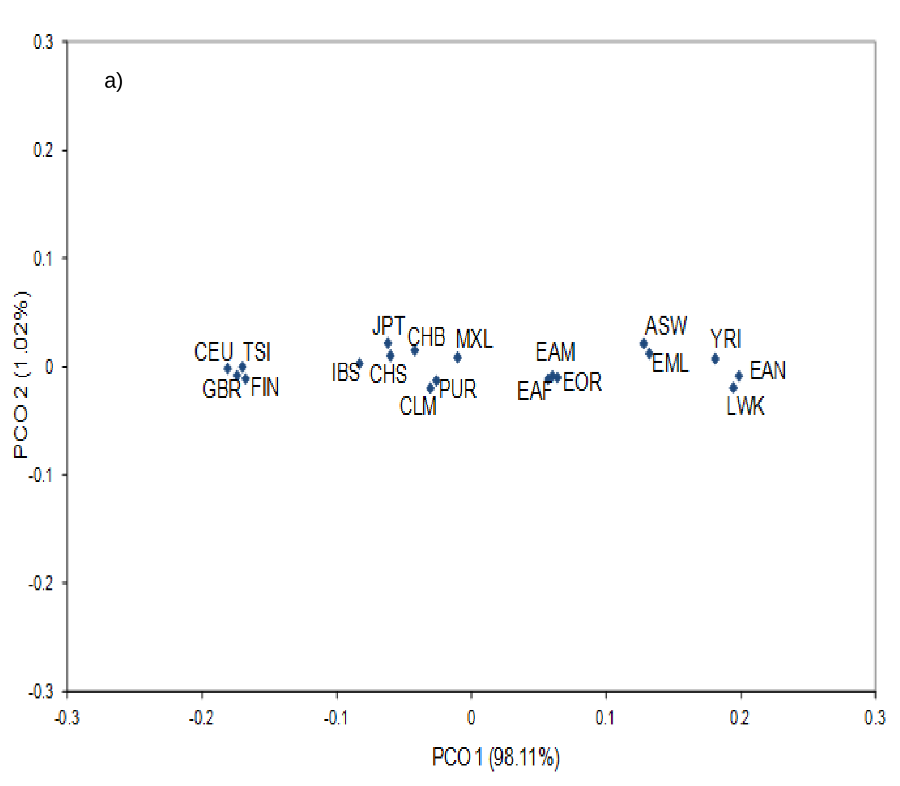
Error bars indicate standard error of the mean

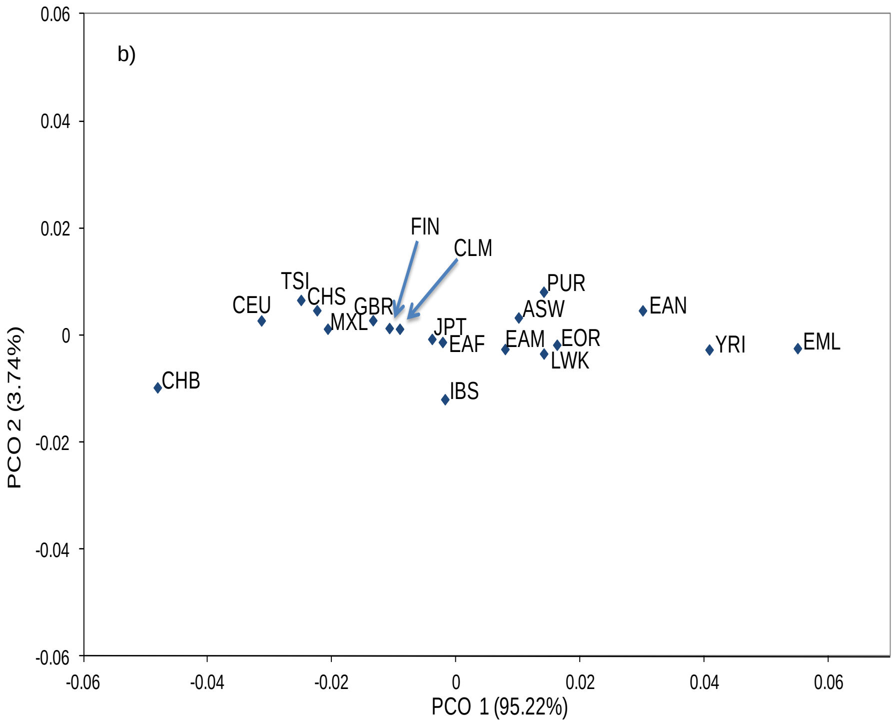
Fig. 4 Gene diversity (h) and nucleotide diversity (π) based on haplotypes inferred for the entire *CYP3A4* gene and *CYP3A4* introns, exons 3' UTR and 5' regulatory region

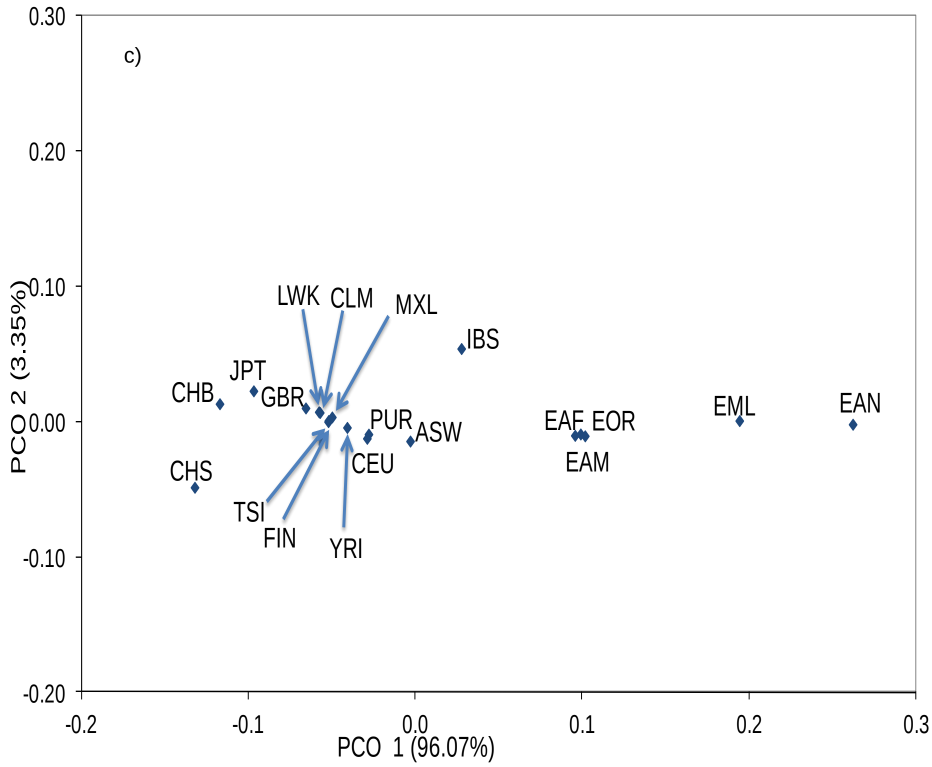


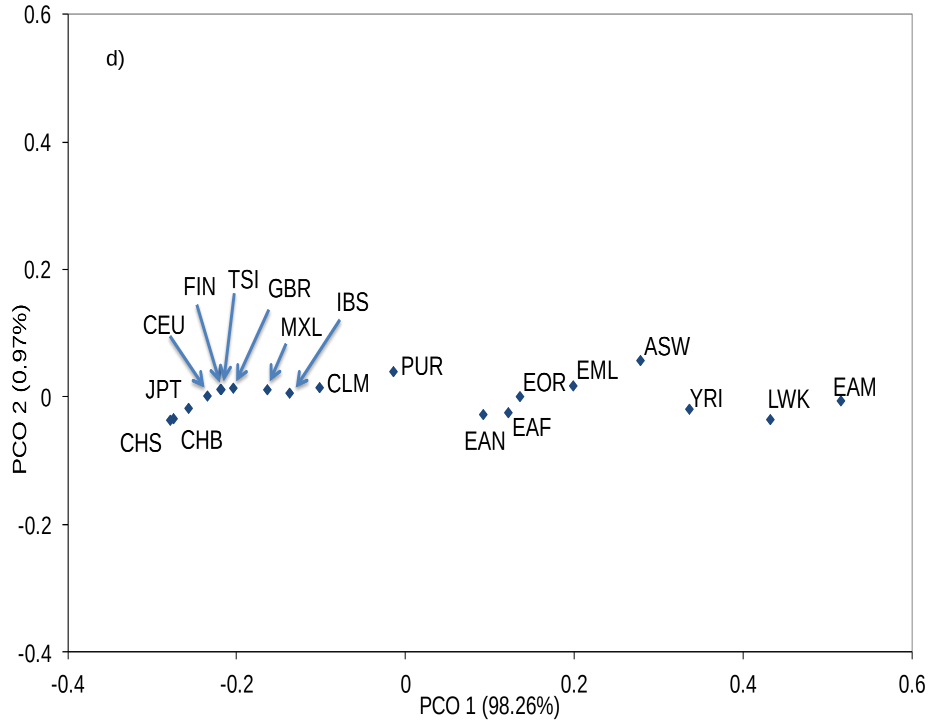
Error bars indicate standard error of the mean. ASW, African American; CEU, European; CHB, Han Chinese; CHS, Southern Han Chinese; CLM, Columbian; FIN, Finnish; GBR, British; IBS, Spanish; JPT, Japanese; LWK, Luhya; MXL, Mexican American; PUR, Puerto Rican; TSI, Tuscan; YRI, Yoruba; EAF, Afar; EAM, Amhara; EAN, Anuak; EML, Maale; EOR, Oromo.

Fig. 5 PCO plots illustrating genetic distances (F_{st}) between populations calculated using: *CYP3A4* a) entire gene, b) coding region, c) 3' UTR and d) 5' regulatory region haplotypes









ASW, African American; CEU, European; CHB, Han Chinese; CHS, Southern Han Chinese; CLM, Columbian; FIN, Finnish; GBR, British; IBS, Spanish; JPT, Japanese; LWK, Luhya; MXL, Mexican American; PUR, Puerto Rican; TSI, Tuscan; YRI, Yoruba; EAF, Afar; EAM, Amhara; EAN, Anuak; EML, Maale; EOR, Oromo. Pairwise F_{st} values are provided in Supplementary Material S3, Figures S2-S5.

Table 1 Frequencies of all *CYP3A4* variants found in Ethiopian groups

CYP3A4 Variant	Chr. position in Human Reference Assembly 36.2	NCBI dbSNP database refSNP ID(s)	Location	Amino acid change	Afar (152)		Amhara (152)		Anuak (152)		Maale (150)		Oromo (146)		Ethiopia Total (752)	
					n	f	n	f	n	f	n	f	n	f	n	f
26897A>T	99354808	28371763	3' UTR		5	0.033	4	0.026	0	0.000	8	0.053	5	0.034	22	0.029
26841C>A	99354864		3' UTR		0	0.000	1	0.007	4	0.026	3	0.020	1	0.007	9	0.012
26716_delT	99354989	28969391	3' UTR		31	0.204	33	0.217	62	0.408	52	0.347	28	0.192	206	0.274
26633G>A	99355072	34141651	3' UTR		0	0.000	2	0.013	0	0.000	1	0.007	1	0.007	4	0.005
26632C>T	99355073	28988604	3' UTR		3	0.020	4	0.026	9	0.059	6	0.040	3	0.021	25	0.033
26537C>T	99355168		3' UTR		0	0.000	0	0.000	0	0.000	1	0.007	0	0.000	1	0.001
26335T>C	99355372		3' UTR		0	0.000	3	0.020	0	0.000	6	0.040	4	0.027	13	0.017
26184C>T	*99355522*		*3' UTR*		*0*	*0.000*	*3*	*0.020*	*0*	*0.000*	*0*	*0.000*	*0*	*0.000*	*3*	*0.004*
26164T>C	*99355541*		*3' UTR*		*0*	*0.000*	*0*	*0.000*	*0*	*0.000*	*4*	*0.027*	*0*	*0.000*	*4*	*0.005*
26076A>T	99355629	35494189	3' UTR		0	0.000	0	0.000	1	0.007	0	0.000	0	0.000	1	0.001
25920T>A	99355785		Exon 13	S495T	0	0.000	1	0.007	0	0.000	0	0.000	2	0.014	3	0.004
25916T>G	99355789		Exon 13		2	0.013	2	0.013	0	0.000	0	0.000	4	0.027	8	0.011
25748C>T	99355957	147972695	Intron 12		2	0.013	1	0.007	5	0.033	1	0.007	3	0.021	12	0.016
25730A>G	99355975	3735451	Intron 12		71	0.467	69	0.454	111	0.730	93	0.620	58	0.397	402	0.535
23090C>T	99358615	12721620	Intron 11		9	0.059	13	0.086	36	0.237	23	0.153	14	0.096	95	0.126
22046A>G	99359659		Intron 11		0	0.000	0	0.000	0	0.000	0	0.000	1	0.007	1	0.001
21971A>G	*99359734*		*Exon 11*	*M395V*	*0*	*0.000*	*0*	*0.000*	*0*	*0.000*	*2*	*0.013*	*0*	*0.000*	*2*	*0.003*
20409C>T	*99361296*		*Intron 10*		*0*	*0.000*	*1*	*0.020*	*0*	*0.000*	*0*	*0.000*	*0*	*0.000*	*1*	*0.001*
20336T>C	99361369	34738177	Intron 10		5	0.033	1	0.007	2	0.013	4	0.027	4	0.027	16	0.021
20318G>C	99361387	4986911	Intron 10		2	0.013	1	0.007	9	0.059	1	0.007	4	0.027	17	0.023
20239G>A (*1G)	99361466	2242480	Intron 10		66	0.434	65	0.428	112	0.737	83	0.553	59	0.404	385	0.512
17974A>G	99363731	34314536	Intron 9		2	0.013	1	0.007	6	0.039	1	0.007	3	0.021	13	0.017
17892C>G	99363813	10267228	Intron 9		29	0.191	28	0.184	49	0.322	29	0.193	24	0.164	159	0.211
*17824_17825delAT (*1S)*	*99363880*	*56153749*	*Intron 9*		*0*	*0.000*	*0*	*0.000*	*2*	*0.013*	*0*	*0.000*	*0*	*0.000*	*2*	*0.003*
17712A>G	99363993		Intron 9		0	0.000	0	0.000	0	0.000	1	0.007	0	0.000	1	0.001
17033C>T	*99364672*	*12721624*	*Intron 8*		*0*	*0.000*	*0*	*0.000*	*2*	*0.013*	*0*	*0.000*	*0*	*0.000*	*2*	*0.003*
16883G>T	99364882		Exon 8	E242STOP	1	0.007	0	0.000	0	0.000	0	0.000	0	0.000	1	0.001
16622C>T	99365083	4646437	Intron 7		67	0.441	69	0.454	114	0.750	91	0.607	62	0.425	403	0.536
15792T>C	99365913	4987160	Intron 7		0	0.000	2	0.013	5	0.033	0	0.000	2	0.014	9	0.012
15762G>T	99365943	2687116	Intron 7		91	0.599	98	0.645	29	0.191	58	0.387	88	0.603	364	0.484
15757A>C	99365948		Intron 7		0	0.000	0	0.000	1	0.007	1	0.007	0	0.000	2	0.003
15658A>T	99366048		Exon 7	Q200H	0	0.000	2	0.013	0	0.000	0	0.000	2	0.014	4	0.005
15637C>T	99366068	4987159	Exon 7		2	0.013	3	0.020	11	0.072	13	0.087	3	0.021	32	0.043
14366T>G	99367339	12721623	Intron 6		1	0.007	0	0.000	0	0.000	1	0.007	0	0.000	2	0.003
14278G>A (*15A)	99367427	4986907	Exon 6	R162Q	0	0.000	2	0.013	1	0.007	0	0.000	1	0.007	4	0.005
14277C>T	99367428	57409622	Exon 6	R162W	0	0.000	0	0.000	0	0.000	0	0.000	1	0.007	1	0.001
14228G>C	99367477		Exon 6	M145I	0	0.000	1	0.007	0	0.000	0	0.000	0	0.000	1	0.001
13978 G>A	99367727		Intron 5		0	0.000	0	0.000	0	0.000	2	0.013	1	0.007	3	0.004
13970 T>C	99367735		Intron 5		1	0.007	0	0.000	0	0.000	0	0.000	0	0.000	1	0.001
13838 G>C	*99367867*	*12721618*	*Intron 4*		*0*	*0.000*	*0*	*0.000*	*3*	*0.020*	*0*	*0.000*	*0*	*0.000*	*3*	*0.004*
6158 T>A	99375547	12721619	Intron 3		1	0.007	0	0.000	2	0.013	4	0.027	1	0.007	8	0.011
4064 T>G	99377641		Exon 2	L47V	0	0.000	1	0.007	0	0.000	0	0.000	0	0.000	1	0.001
3873_3875delATT	99377827:99363881	71581993	Intron 1		0	0.000	0	0.000	0	0.000	4	0.027	1	0.007	5	0.007
144 A>G	99381561	35587942	Intron 1		0	0.000	0	0.000	1	0.007	4	0.027	1	0.007	6	0.008
102 T>C	99381603		Intron 1		7	0.046	2	0.013	5	0.033	1	0.007	8	0.055	23	0.031
66 C>G	99381639		Exon 1		1	0.007	0	0.000	0	0.000	3	0.020	1	0.007	5	0.007
-32 T>C	99381736		5'upstream		0	0.000	0	0.000	0	0.000	0	0.000	1	0.007	1	0.001
-62 C>A	99381766	12721636	5'upstream		9	0.059	3	0.020	0	0.000	4	0.027	10	0.068	26	0.035
-247G>A	99381952		5'upstream		0	0.000	0	0.000	0	0.000	0	0.000	1	0.007	1	0.001
-333C>T	*99381998*		*5'upstream*		*0*	*0.000*	*0*	*0.000*	*0*	*0.000*	*2*	*0.013*	*0*	*0.000*	*2*	*0.003*
-392A>G	99382096	2740574	5'upstream		53	0.349	49	0.322	125	0.822	71	0.473	52	0.356	350	0.465
-7949G>A	99389685		5'upstream		0	0.000	0	0.000	1	0.007	0	0.000	0	0.000	1	0.001
-10502C>T	99392236	115518242	5'upstream		1	0.007	1	0.007	4	0.026	0	0.000	1	0.007	7	0.009
-10755_delG	99392451		5'upstream		2	0.013	0	0.000	0	0.000	1	0.007	1	0.007	4	0.005
-10852C>T	99392586	74370707	5'upstream		0	0.000	0	0.000	2	0.013	1	0.007	0	0.000	3	0.004
-11070G>C	99392804	36015827	5'upstream		0	0.000	2	0.013	1	0.007	0	0.000	0	0.000	3	0.004
-11073T>C	99392807		5'upstream		0	0.000	0	0.000	0	0.000	1	0.007	0	0.000	1	0.001
-11131G>A	*99392865*	*112639771*	*5'upstream*		*0*	*0.000*	*0*	*0.000*	*4*	*0.026*	*0*	*0.000*	*0*	*0.000*	*4*	*0.005*
-11185_-11186insTGT	99392921:99392922	34401238	5'upstream		3	0.020	5	0.033	0	0.000	0	0.000	0	0.000	8	0.011
-11386T>A	99393120	34020359	5'upstream		0	0.000	2	0.013	5	0.033	3	0.020	0	0.000	10	0.013

Numbers in brackets indicate the number of chromosomes. Italics indicate variants found on more than one chromosome that were private to a population. Shaded rows indicate variants found both in Ethiopia and populations of The 1000 Genomes Project. n, the number of chromosomes carrying the variant allele. f, frequency of the variant allele.

Table 3 The predicted effect of *CYP3A4* non synonymous variants on protein structure and function as determined by PolyPhen2

***CYP3A4* Variant**	**Exon**	**Amino acid change**	**Frequency**	**Predicted effect on CYP3A4 structure/function**
6003G>A	3	G56D	0.005 FIN, 0.005 TSI	Probably damaging
14278G>A	6	R162Q	0.007 EAN, 0.007 EOR, 0.008 ASW, 0.008 MXL, 0.012 EAM, 0.028 YRI	Benign
14313G>C	6	D174H	0.009 PUR, 0.017 GBR	Benign
15612C>G	7	T185S	0.022 JPT	Possibly damaging
15658A>T	7	Q200H	0.005 CHS, 0.011 JPT, 0.013 EAM, 0.014 EOR	Benign
15722T>C	7	S222P	0.011 FIN	Possibly damaging
20079T>C	10	L293P	0.015 CHS, 0.017 JPT	Benign
21905C>T	11	L373F	0.026 LWK	Benign
21971A>G	11	M395V	0.013 EAM	Benign
23181T>C	12	M445T	0.006 CEU, 0.006 GBR, 0.008 CLM, 0.011 FIN	Probably damaging
25920T>A	13	S495T	0.007 EAM, 0.014 EOR	Benign

ASW, African American; CEU, European; CHS, Southern Han Chinese; FIN, Finnish; GBR, British; JPT, Japanese; LWK, Luhya; MXL, Mexican American; PUR, Puerto Rican; TSI, Tuscan; YRI, Yoruba; EAM, Amhara; EAN, Anuak; EML, Maale; EOR, Oromo

Table 4 Frequencies of *CYP3A4* coding region haplotypes

	Population	ASW	CEU	CHB	CHS	CLM	FIN	GBR	IBS	JPT	LWK	MXL	PUR	TSI	YRI	EAF	EAM	EAN	EML	EOR
Haplotype ID	1	0.951	0.994	1.000	0.985	0.983	0.973	0.978	1.000	0.961	0.928	0.992	0.945	0.990	0.892	0.967	0.934	0.921	0.880	0.911
	2															0.007			0.020	0.007
	3						0.005							0.005						
	4	0.008										0.008			0.028		0.013	0.007		0.007
	5							0.017					0.009							
	6									0.022										
	7	0.041				0.008					0.046		0.045	0.005	0.080	0.013	0.020	0.072	0.087	0.021
	8																0.013			0.014
	9						0.011													
	10				0.010					0.006										
	11				0.005					0.011										
	12										0.026									
	13																		0.013	
	14		0.006			0.008	0.011	0.006												
	15															0.013	0.013			0.027
	16																0.007			0.014

ASW, African American; CEU, European; CHB, Han Chinese; CHS, Southern Han Chinese; CLM, Columbian; FIN, Finnish; GBR, British; IBS, Spanish; JPT, Japanese; LWK, Luhya; MXL, Mexican American; PUR, Puerto Rican; TSI, Tuscan; YRI, Yoruba; EAF, Afar; EAM, Amhara; EAN, Anuak; EML, Maale; EOR, Oromo.

Table 5 The predicted impact of *CYP3A4* variants on consensus splice sites

***CYP3A4* variant**	**Location**	**Distance from exon (nucleotides)**	**Distribution in World Dataset**	**Impact on consensus splice site[1]**	**Predicted impact on splicing[1]**
25835A	Intron 12	19	0.015 LWK	None	None
23090T	Intron 11	11	0.008 MXL; 0.025 CLM; 0.059 EAF; 0.073 PUR; 0.086 EAM; 0.096 EOR; 0.153 EML; 0.232 LWK; 0.237 EAN; 0.295 ASW; 0.330 YRI	None	None
22050T	Intron 11	9	0.011 GBR;0.008 MXL	None	+7 nucleotides added to exon 11 due to creation of putative alternative splice site
20239A	Intron 10	12	0.705 ASW;0.052 CEU; 0.247 CHB; 0.245 CH; 0.292 CLM; 0.098; 0.434 EAF; 0.428 EAM; 0.737 EAN; 0.553 EML; 0.404 EOR; 0.070 FIN; 0.079 GBR; 0.143 IBS; 0.281 JPT; 0.997 LWK; 0.417 MXL; 0.318 PUR; 0.082 TSI; 0.858 YRI	Possible disruption	+11 nucleotides added to exon 10 due to creation of putative alternative splice site
14319A	Intron 6	5	0.025 ASW; 0.015 LWK; 0.009 PUR; 0.011 YRI	Possible disruption	None
14209G	Intron 5	17	0.005 CHB	None	None
13970G	Intron 5	10	0.007 EAF	None	None
13838C	Intron 4	9	0.020 EAN; 0.010 LWK; 0.009 PUR; 0.023 YRI	None	None
11975C	Intron 3	18	0.005 CHS	None	None

ASW, African American; CEU, European; CHB, Han Chinese; CHS, Southern Han Chinese; CLM, Columbian; FIN, Finnish; GBR, British; IBS, Spanish; JPT, Japanese; LWK, Luhya; MXL, Mexican American; PUR, Puerto Rican; TSI, Tuscan; YRI, Yoruba; EAF, Afar; EAM, Amhara; EAN, Anuak; EML, Maale; EOR, Oromo. [1] Predictions made using Human Splicing Finder Software.

Table 6 The predicted impact of *CYP3A4* 5' regulatory region variants on transcription factor binding sites

***CYP3A4* variant**	**Distribution in Ethiopia**	**Location**	**Possible Implications for *CYP3A4* expression**		
			Predicted effect on transcription factor binding sites[1]	**Additional evidence**	**Putative expression phenotype**
-333T	EML 1.3%	Promoter	Disruption of putative GRE	Hypothesised that a GRE exists within the promoter (El Sankary et al., 2002)	Reduced expression
-392A	EAN, EOR, EML, EAF, EAM ≤ 82%	Promoter	Creation of putative PPRE	Increased and decreased expression observed *in vitro* (Wandel et al., 2000; Spurdle et al., 2002; Amirimani et al., 2003; Rodríguez-Antona et al., 2005)	Increased expression
-10502C	EAN, EOR, EAF, EAM ≤2.6%	Downstream of CLEM4	Creation of putative C/EBP site	Evidence that C/EBPα up regulates *CYP3A4* by direct binding at promoter (Rodríguez-Antona et al., 2003; Bombail et al., 2004).	Increased expression
-11131A	EAN 2.6%	CLEM4	Disruption of functional Ebox	Site directed mutagenesis of this E-box reported to reduce enhancer activity by 43% *in vitro* (Matsumara et al., 2004)	Reduced expression
-11185insTGT	EAM, EAF ≤3.3%	CLEM4	Disruption of functional Ebox	-11185_-11186insTGT reported to reduce enhancer activity *in vitro* (Matsumara et al., 2004)	Reduced expression

EAF, Afar; EAM, Amhara; EAN, Anuak; EML, Maale; EOR, Oromo; GRE, Glucocorticoid Response Element; PPRE, Peroxisome Proliferator Response Element; C/EBP, CCAAT Enhancer Binding Protein. [1] Predictions made using MATCH and Matinspector.

Table 7 Frequencies of *CYP3A4* 5' regulatory region haplotypes within Ethiopia

Nucleotide change[1]	-11185_-11186_insTGT	-11386 T>A	-11131 G>A	-11073 T>C	-11070 T>C	-10852 C>T	-10755 _delG	-10502 C>T	-7949 G>A	-392 A>G	-333 C>T	-247 G>A	-62 C>A	-32 T>C	Haplotype frequencies EAM (152)	EAN (152)	EOR (146)	EAF (152)	EML (150)	Ethiopia Pooled (752)
Location	CLEM4								XREM	Promoter										
Haplotype ID[2] 1															0.625	0.143	0.570	0.566	0.508	0.482
2															0.023	0.000	0.070	0.057	0.025	0.035
3															0.289	0.722	0.313	0.336	0.402	0.412
4															0.000	0.000	0.016	0.000	0.000	0.003
5															0.000	0.000	0.016	0.000	0.000	0.003
6															0.000	0.000	0.000	0.000	0.016	0.003
7															0.000	0.008	0.000	0.000	0.000	0.002
8															0.008	0.032	0.008	0.008	0.000	0.011
9															0.000	0.000	0.008	0.008	0.008	0.005
10															0.000	0.016	0.000	0.000	0.008	0.005
11															0.008	0.008	0.000	0.000	0.000	0.003
12															0.000	0.000	0.000	0.000	0.008	0.002
13															0.000	0.032	0.000	0.000	0.000	0.006
14															0.008	0.040	0.000	0.000	0.025	0.014
15															0.039	0.000	0.000	0.025	0.000	0.013

Numbers in brackets indicate number of chromosomes. EAF, Afar; EAM, Amhara; EAN, Anuak; EML, Maale; EOR, Oromo;

[1] Position from base A in initiation codon (A in ATG is considered as +1)

[2] White cell, allele observed in *CYP3A4*1A;* grey cell, derived allele.

Table 8 Frequencies of variants found upstream of *CYP3A7*

Region	Country	n	Frequency of *CYP3A7* variant				
			*CYP3A7*1D*	-91A	*CYP3A7*1C*	-239C	*CYP3A7*1B*
North Africa	**Algeria**						
	Algerian	262	0.034	0.000	0.073	0.000	0.000
	Morocco						
	Ifrane Berber	160	0.031	0.000	0.175	0.000	0.000
	Moroccan Jews	176	0.011	0.000	0.028	0.000	0.000
Central and East Africa	**Ethiopia**						
	Afar	144	0.139	0.000	0.042	0.000	0.000
	Amhara	144	0.09	0.000	0.063	0.014	0.000
	Anuak	146	0.164	0.000	0.007	0.014	0.000
	Maale	142	0.106	0.000	0.056	0.000	0.000
	Oromo	136	0.059	0.000	0.022	0.000	0.000
	Sudan						
	Various	58	0.103	0.000	0.017	0.017	0.000
	North	192	0.068	0.000	0.021	0.000	0.000
	South	208	0.216	0.000	0.000	0.014	0.000
West Africa	**Cameroon**						
	Lake Chad	230	0.213	0.000	0.026	0.009	0.000
	Mayo Darle	214	0.215	0.000	0.033	0.005	0.000
	Mambila	130	0.077	0.000	0.031	0.000	0.000
	Ghana						
	Asante	66	0.121	0.000	0.000	0.000	0.000
	Bulsa	144	0.097	0.000	0.000	0.000	0.000
	Kasena	54	0.056	0.000	0.000	0.000	0.000
	Nigeria						
	Igbo	158	0.101	0.000	0.000	0.000	0.000
	Senegal						
	Manjak	120	0.158	0.000	0.000	0.017	0.000
	Wolof	126	0.135	0.000	0.016	0.032	0.000
Central Africa	**Congo**						
	Brazzaville	102	0.108	0.000	0.000	0.000	0.000
South East Africa	**Malawi**						
	Chewa	178	0.089	0.000	0.006	0.000	0.000
	Tumbuka	112	0.161	0.000	0.000	0.000	0.000
	Yao	106	0.085	0.000	0.000	0.000	0.000
	Various	92	0.141	0.000	0.000	0.012	0.000
	Mozambique						
	Sena	122	0.180	0.000	0.000	0.000	0.000
	Various	66	0.197	0.000	0.000	0.000	0.000
	South Africa						
	Bantu Speakers	74	0.135	0.000	0.000	0.000	0.000
	Tanzania						
	Chagga	96	0.135	0.000	0.010	0.000	0.000
	Uganda						
	Bantu Speakers	78	0.154	0.000	0.000	0.000	0.000

	Zimbabwe						
	Shona	82	0.085	0.000	0.012	0.000	0.000
Middle East	**Israel**						
	Sephardi Jews	124	0.000	0.008	0.000	0.000	0.000
	Yemen						
	Hadramaut	152	0.020	0.013	0.013	0.000	0.000
	Msila and Sena	74	0.122	0.000	0.000	0.000	0.000
Europe	**Armenia**						
	Armenian	142	0.000	0.000	0.028	0.000	0.000
	Netherlands						
	Friesians	166	0.000	0.000	0.036	0.000	0.006
	Turkey						
	Anatolian	130	0.000	0.008	0.008	0.000	0.000
	United Kingdom						
	British						
		162	0.000	0.000	0.043	0.000	0.006
	Ukraine						
	Ashkenazi Jews	166	0.000	0.012	0.012	0.000	0.000

n represents the number of chromosomes.

Table 9 Allele frequencies of *CYP3A7*1C, CYP3A5*3, CYP3A5*6, CYP3A5*7* in Algerians and Moroccan Berbers.

Population	***CYP3A7*1C***	***CYP3A5*3***	***CYP3A5*6***	***CYP3A5*7***
Algerians (254)	0.075	0.862	0.047	0.008
Moroccan Berbers (156)	0.179	0.859	0.038	0.000

Numbers in brackets indicate the number of chromosomes.

Table 10 *CYP3A7/CYP3A5* haplotypes inferred within Algerians and Moroccan Berbers.

***CYP3A* variant**		***CYP3A7*1C***	***CYP3A5*6***	***CYP3A5*3***	***CYP3A5*7***	**Haplotype frequency**	
						Algerians (254)	**Moroccan Berbers (156)**
Haplotype ID[1]	**1**					0.126	0.141
	2					0.008	0.000
	3					0.744	0.641
	4					0.047	0.038
	5					0.004	0.000
	6					0.071	0.179

Numbers in brackets indicate the number of chromosomes.

[1] White cell, allele observed in *CYP3A7*1A* or *CYP3A5*1A*; grey cell, derived allele.

Table 11 Putative phenotypes for haplotypes containing *CYP3A7*1C* and *CYP3A4* regulation variants.

Haplotype	Variants defining haplotype	Hypothesised Expression	Haplotype Frequency				
			Afar (120)	Amhara (120)	Anuak (118)	Maale (120)	Oromo (112)
13	*CYP3A7*1C* and *CYP3A4*1A*	↑ CYP3A7	0.025	0.067	0.008	0.042	0.009
14	*CYP3A7*1C* and *CYP3A4* -333T	↑ CYP3A7 ↓ CYP3A4	0.000	0.000	0.000	0.017	0.000
15	*CYP3A7*1C* and *CYP3A4*1B*	↑ CYP3A7 ↑ CYP3A4	0.000	0.000	0.000	0.000	0.009

Numbers in brackets indicate the number of chromosomes. Upward arrows indicate a putative increase in expression associated with haplotype, downward arrows indicate a putative decrease in expression associated with haplotype.

Table 12 Location of putative transcription factor binding sites predicted to be created/destroyed/unchanged by *CYP3A7*1D*.

Transcription Factor	Prediction Tool	Position	CSS	MSS	Predicted effect of *CYP3A7*1D*
GRE	MATCH	-53/-58	0.97	0.92	Site created
AHR	MATCH	-55/-47	1.00	0.98	Site destroyed
C/EBP	MATCH	-51/-34	0.94	0.91	Site unchanged

GRE, glucocorticoid response element; AHR, aryl hydrocarbon receptor binding site; C/EBP, CCAAT enhancer binding protein site; CSS, core similarity score; MSS, matrix similarity score.

Supplementary Digital Content

Supplementary Material S1

Table S1 Details of samples used for the geographic survey of variants associated with adult expression of CYP3A7

Country	N	Cultural ID	Collection Location
Algeria	131	Not specified	Port Say, Mostaganem, Oran
Armenia	71	Not specified	North and South Armenia
Cameroon	107 115 65	Various Various Mambila	Mayo Darle Lake Chad Somie Grassfields
Congo	51	Not specified	Brazzaville
Ghana	33 72 27	Asante Bulsa Kasena	Enchi Sandema Navrongo
Israel	62	Sephardi Jew	Israel
Malawi	89 56 53 46	Chewa Tumbuka Yao Various	All collected from: Lilongwe, Kanengo, Mzuzu and Mangochi
Morocco	80 88	Ifrane Berber Morrocan Jews	Ifrane
Mozambique	61 33	Various Various	Sena Duwa, Tembo, Malunga, Zora, Chirengi
Netherlands	83	Friesian	Friesland
Nigeria	79	Igbo	Calabar
Senegal	60 63	Manjak Wolof	South Dakar
South Africa	37	Bantu speakers	Pretoria
Sudan	29 96 104	Various North Sudanese South Sudanese	Ommrawaba North South
Tanzania	48	Chagga	Kilimanjaro
Turkey	65	Anatolian Turks	West and East Anatolia
Uganda	39	Bantu speakers	Ssese
Ukraine	83	Ashkenazi Jews	Diaspora and Israel
United Kingdom	81	British	Chippenham and North Walsham
Yemen	37 76	Unspecified Unspecified	Sena and Msila Hadramaut
Zimbabwe	41	Shona	Mposi

N represents the number of individuals.

Supplementary Material S2

Amplification and sequencing *CYP3A4* exons 1 to 7

Primers designed to amplify *CYP3A4* exons 1 to 7 and regions of flanking adjacent introns are provided in Table S2.

DNA was amplified within 96-well plates in 10 µl reaction volumes containing 1 µl (~ 1 ng) of template DNA, 0.5 µM of each oligonucleotide, 0.2 µM dNTPs, 0.2 units Taq DNA polymerase (HT Biotechnology, Cambridge, UK) and 1x reaction buffer (10 mM Tris-HCl, pH 9.0, 1.5 mM $MgCl_2$, 50 mM KCl, 0.1% Triton X-100, 0.01% gelatin). The cycling parameters for amplification were: an initial denaturation step of 5 minutes at 94ºC, followed by 35 cycles of denaturation at 94ºC for 30 seconds, annealing at the temperatures provided in Table S2 for 30 seconds and extension at 72ºC for 30 seconds. A final extension step at 72ºC for 10 minutes completed the cycle.

The PCR product was purified by mixing with 30 µl of 1/3 water 2/3 home-made micro clean (HM-MC) (40 % PEG-8000, 1 M NaCl, 2 mM Tris-HCl (pH 7.5), 0.2 mM EDTA, 3.5 mM $MgCl_2$), followed by centrifugation at 4000 g for 60 minutes, removal of the supernatant and addition of 150 µl of 70% ethanol. This was followed by further centrifugation at 4000 g for 25 min and removal of the supernatant. Samples were dried at 65 ºC for 5 minutes and resuspended in 10 µl distilled water. This purified PCR product was used as the DNA template for sequencing reactions.

Bidirectional sequencing of purified PCR products was conducted using the primers detailed in Table S2. DNA was sequenced in 96-well plates in 10 µl reaction volumes containing 6 µl of the purified PCR product, 0.32 µl of either primer (at 5 µM), 0.32 µl of BigDye termination mix v3.1 (BigDye® Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems (ABI), Warrington UK) and 2.15 µl of Better Buffer (Applied Biosystems (ABI), Warrington UK). The cycling parameters were: an initial step of 96 ºC for 1 minute followed by 25 cycles of 96 ºC for 10 seconds, 55 ºC for 10 seconds and 60 ºC for 4 minutes.

Sequencing products were purified by ethanol precipitation. To each 10 µl reaction 2.5 µl of 125 mM EDTA was added, followed by 30 µl of 100% ethanol and centrifugation at 4000 g for 60 minutes. The supernatant was removed and a further 30 µl of 70% ethanol added followed by centrifugation for 10 minutes and removal of the supernatant. Samples were dried at 65 ºC for 5 minutes. Each sample was subsequently mixed with 10 µl of high purity formamide, heated at 96 ºC for five minutes, cooled on ice and subsequently run on an ABI 3730 genetic analyser. Sequence traces were analysed using Sequencher 4.1.10 software (Gene Codes Corporation, USA).

Amplification and sequencing of *CYP3A4* exons 8 to 13, the 5' regulatory region and 3' UTR

CYP3A4 exons 8 to 13 (and regions of flanking adjacent introns), the 3' UTR and 5' regulatory regions were initially sequenced by Macrogen (www.Macrogen.com). 3.2 ng of genomic DNA template was sent to Macrogen for PCR amplification and sequencing in 96-well plates. Sequencing conducted by Macrogen was in the forward direction only. Sequence traces were analysed using Sequencher 4.1.10 software (Gene Codes Corporation, USA).

Primers used for PCR amplification and sequencing of exons 8 to 13 and the 3' UTR are provided in Table S2. PCR amplification and sequencing (bi-directional sequencing) was repeated for all identifications of novel variation using the methods described for exons 1 to 7.

The 5' regulatory regions sequenced comprised the *CYP3A4* promoter and the two upstream *CYP3A4* enhancers: the xenobiotic responsive enhancer module (XREM) and the constitutive liver enhancer module 4 (CLEM4). These regions are highly conserved functional DNA sequences containing clusters of binding sites for transcription factors that regulate *CYP3A4*. This upstream arrangement is unique to the *CYP3A* family (Qiu et al., 2010).

The promoter (-500 nucleotides), XREM (-7.7 kb/-7.9 kb) and CLEM4 (-10.9 kb/ -11.4 kb) were sequenced using primers detailed in Table S2.

Table S2 Primers for PCR amplification and sequencing of the *CYP3A4* promoter, upstream enhancers, exons (and flanking regions of adjacent introns) and the 3' UTR.

Region Amplified	Primer ID	Primer Sequence 5'- 3'	Amplicon Length/nt	PCR Annealing temperature/°C
Promoter	Forward Reverse	CTCTGTCTGTCTGGGTTTGG CCTTTCAGCTCTGTGTTGCTC	501	56
XREM	PCR Forward PCR Reverse Seq. Forward Seq. Reverse	GCAAAGGTCTTGTTCACACC TAGTGTTTACCATGTGCACAT GTCTCTCTGGGGTCCCCT TAGTGTTTACCATGTGCACAT	669 440	56
Part one of CLEM4	Forward Reverse	TGGCAGGCACTGGAATTG CTCTGAGCCACAGAGTGACC	562	56
Part two of sCLEM4 And flanking region	Forward Reverse	TGGCTGTTTCCCACCTTC GGACTCCAGAGACATCTCGTTC	547	56
Exon 1*	PCR Forward PCR Reverse Seq. Forward Seq. Reverse	CAGGCATAGGTAAAGATCTGTAGGTG ACCTCAGGCAGTCCACTTGC AGGTCAGTGAGTGGTGTGTGT ACCTCAGGCAGTCCACTTGC	722	62
Exon 2	Forward Reverse	GCCACTTACTAAATAGACGTAAAGGAACA TGGGGTAAACATTGCCATGT	603	61
Exon 3	Forward Reverse	CTCTTATTTCCTTTATGACGTCTCC CGAAGGATAATTATTTCAAAACTGATAA	498	59
Exon 4	Forward Reverse	TCCTTCATCATATGAAGACTTGA AGAACAGAAATGAGGGCACATAAAG	572	52.5
Exon 5*	PCR Forward PCR Reverse Seq. Forward Seq. Reverse	AGCATAGGGCCCATCACCC TGGATATGTAAACCCTGGCCC AGCATAGGGCCCATCACCC AGGACATGGCTTTCCCCAGCAT	832	63
Exon 6*	PCR Forward PCR Reverse Seq. Forward Seq. Reverse	AGCATAGGGCCCATCACCC TGGATATGTAAACCCTGGCCC GAATGAATCTGGTGGGGACAGG TGGATATGTAAACCCTGGCCC	832	63
Exon 7	Forward Reverse	TGGAGTGTGATAGAAGGTGATCT CTGATAGCTAAAAATGTATGAGGTC	503	56
Exon 8	Forward Reverse	TCTAGGAGACTGTAGTCCAATAG GAGCAGTCTTCATGTTAAAAGCA	561	56
Exon 9	Forward Reverse	AGGGTATGTTTTTCACTGGTGAT CCCACAATTAATTTTGCAG	577	56
Exon 10	Forward Reverse	ATGAAACCACCCCCAGTGTAC AGTAATAGAAAGCAGATGAACCAGA	540	56
Exon 11	Forward Reverse	TCGATCCTTTACCAGTATGAGTTAG TTGGAATTGTGGATGACTG	558	56
Exon 12	Forward Reverse	GTGTCAGGAGAGTAGAAAGGATCTGTAG CCATGCTAATCTACATGGGCTTTA	581	56
Exon 13 (part 1)	Forward Reverse	CTCTCACTGTCCAATCTTCACA GAGCCAAATCTACCTCCTCAC	587	56
Exon 13 (part 2)	Forward Reverse	TGGGCTTCATCCAATGGA GCTCAGCCTCCCAAACTGCTA	618	56
Exon 13 (part 3) and UTR	Forward Reverse	AAGTTAATCCACTGTGACTTTG GCTTTAGGACTAAATTATTCAGG	649	56

* denotes exons for which a different set of primers were used during sequencing.

Supplementary Material S3

Figure S1 *CYP3A4* haplotypes inferred using the world dataset

Frequencies of haplotypes are included in Table S3. Blue cell, presence of non-*CYP3A4*1A* allele; white cell, allele observed in *CYP3A4*1A;* red cell, non synonymous variant predicted to impact protein structure/function; grey cell, putative splice variant.

Table S3 Frequencies of *CYP3A4* entire gene haplotypes within individual populations

Haplotype ID	ASW	CEPH	CHB	CHS	CLM	FIN	GBR	IBS	JPT	LWK	MXL	PUR	TSI	YRI	EAF	EAN	EAM	EML	EOR
1	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
2	0.06	0.00	0.00	0.00	0.06	0.03	0.03	0.00	0.00	0.16	0.00	0.01	0.02	0.09	0.01	0.00	0.01	0.01	0.02
3	0.11	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.02	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00
4	0.01	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
7	0.05	0.03	0.12	0.10	0.00	0.03	0.04	0.07	0.08	0.06	0.14	0.03	0.05	0.07	0.00	0.00	0.00	0.01	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
13	0.01	0.00	0.11	0.12	0.05	0.00	0.00	0.00	0.00	0.01	0.07	0.06	0.00	0.01	0.03	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.02
16	0.04	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.16	0.01	0.02	0.00	0.07	0.14	0.13	0.11	0.05	0.07
17	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00
18	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00
21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.00	0.00	0.00
22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
23	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
24	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.01	0.03	0.01	0.05	0.00
25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
26	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.03	0.00	0.01	0.00	0.00	0.01
27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00
31	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00
32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
33	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00
35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.02	0.01
37	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00
38	0.00	0.00	0.10	0.07	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.03	0.03	0.01
39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01

42	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00
45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
47	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
48	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
49	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.04	0.00	0.17	0.00	0.01	0.00	0.00	0.00
50	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
51	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
53	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.00	0.00
54	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
57	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
58	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
60	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
61	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
62	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
63	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
64	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.00	0.00
65	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.00	0.00
66	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
68	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
69	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.02	0.00	0.01	0.00	0.01	0.00	0.00	0.00
70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00
71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
72	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
73	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
74	0.00	0.00	0.01	0.03	0.09	0.00	0.00	0.04	0.01	0.00	0.14	0.02	0.00	0.00	0.00	0.00	0.03	0.01	0.01
75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
76	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
77	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
78	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.00
80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00
81	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01
82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
84	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.03	0.04	0.05
85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
86	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.03	0.03	0.03	0.01	0.03

87	0.03	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.03	0.02	0.07	0.01	0.03	0.05
88	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00
89	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
91	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
92	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
93	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00
94	0.20	0.86	0.63	0.67	0.63	0.84	0.85	0.82	0.67	0.05	0.56	0.62	0.84	0.07	0.36	0.05	0.34	0.15	0.32
95	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.03	0.00	0.00
96	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
97	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
98	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.04
100	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
101	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
102	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
103	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
104	0.02	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.01	0.00
105	0.00	0.00	0.00	0.00	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
106	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
107	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
108	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
109	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
110	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
111	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
112	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
113	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
114	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.03	0.01	0.01	0.02
115	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
116	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
117	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
118	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.02
119	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
120	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
121	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
122	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.00
123	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00
124	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
125	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
126	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.02	0.01	0.01	0.00	0.02	0.00	0.00	0.00	0.00	0.00
127	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
128	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
129	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
130	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00
131	0.00	0.02	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00

132	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
133	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
134	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
135	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.07	0.03	0.05	0.01
136	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
137	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.01
138	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
139	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.01	0.06	0.04	0.03
140	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
141	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00
142	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
143	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
144	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00
145	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
146	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
147	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
148	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00
149	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
150	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.13	0.02	0.02	0.03
151	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00
152	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
153	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
154	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
155	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
156	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
157	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
158	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
159	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.01
160	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.02	0.01
161	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
162	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
163	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
164	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
165	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
166	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
167	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
168	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
169	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
170	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
171	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.02	0.02
172	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01
173	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
174	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
175	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
176	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00

177	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
178	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
179	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
180	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.01	0.03	0.00
181	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
182	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
183	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
184	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01
185	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
186	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
187	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
188	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
189	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
190	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
191	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01
192	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
193	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
194	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
195	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
196	0.00	0.02	0.00	0.00	0.00	0.01	0.00	0.04	0.00	0.00	0.00	0.01	0.02	0.00	0.01	0.00	0.01	0.01	0.01
197	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
198	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.00	0.01	0.02	0.02
199	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
200	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
201	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
202	0.03	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
203	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Haplotype IDs shaded in blue are those unique to Ethiopia, those shaded in green are unique to African populations. Population codes: ASW, African American; CEU, European; CHB, Han Chinese; CHS, Southern Han Chinese; CLM, Columbian; FIN, Finnish; GBR, British; IBS, Spanish; JPT, Japanese; LWK, Luhya; MXL, Mexican American; PUR, Puerto Rican; TSI, Tuscan; YRI, Yoruba; EAF, Afar; EAM, Amhara; EAN, Anuak; EML, Maale; EOR, Oromo

Table S4 Nucleotide diversity (π) of *CYP3A4* gene regions for individual populations in ascending order

Region of *CYP3A4*															
3' UTR				5' Regulatory region				Exons				Introns			
Population	π	SD	K	Population	π	SD	K	Population	π	SD	K	Population	π	SD	K
CHS	0.00000	0.000	0.00	IBS	0.00000	0.000	0.00	CHB	0.00000	0.000	0.00	CEPH	0.00013	0.000	0.41
CEPH	0.00001	0.000	0.01	CHS	0.00001	0.000	0.01	IBS	0.00000	0.000	0.00	FIN	0.00017	0.000	0.54
CHB	0.00001	0.000	0.01	JPT	0.00001	0.000	0.01	CEPH	0.00001	0.000	0.01	GBR	0.00018	0.000	0.58
JPT	0.00002	0.000	0.02	CEPH	0.00002	0.000	0.34	MXL	0.00001	0.000	0.02	IBS	0.00020	0.000	0.65
MXL	0.00005	0.000	0.06	CHB	0.00003	0.000	0.06	TSI	0.00001	0.000	0.02	TSI	0.00020	0.000	0.64
CLM	0.00006	0.000	0.07	FIN	0.00003	0.000	0.06	CLM	0.00002	0.000	0.03	CHS	0.00031	0.000	0.99
GBR	0.00006	0.000	0.07	GBR	0.00004	0.000	0.08	CHS	0.00003	0.000	0.04	JPT	0.00033	0.000	1.06
LWK	0.00008	0.000	0.09	TSI	0.00004	0.000	0.07	GBR	0.00003	0.000	0.05	CHB	0.00034	0.000	1.10
FIN	0.00008	0.000	0.10	MXL	0.00009	0.000	0.18	FIN	0.00004	0.000	0.05	MXL	0.00040	0.000	1.42
TSI	0.00009	0.000	0.10	CLM	0.00010	0.000	0.20	EAF	0.00004	0.000	0.07	CLM	0.00045	0.000	1.43
IBS	0.00012	0.000	0.14	PUR	0.00019	0.000	0.37	ASW	0.00006	0.000	0.10	PUR	0.00066	0.000	2.12
PUR	0.00014	0.000	0.16	EAN	0.00026	0.000	0.50	JPT	0.00007	0.000	0.10	LWK	0.00072	0.000	2.31
YRI	0.00020	0.000	0.23	LWK	0.00031	0.000	0.59	PUR	0.00008	0.000	0.11	YRI	0.00078	0.000	2.49
ASW	0.00033	0.000	0.38	EAM	0.00032	0.000	0.61	LWK	0.00009	0.000	0.14	EAM	0.00079	0.000	2.52
EAF	0.00037	0.000	0.43	EOR	0.00032	0.000	0.62	EAM	0.00009	0.000	0.13	EAF	0.00083	0.000	2.66
EOR	0.00043	0.000	0.50	EAF	0.00034	0.000	0.65	EAN	0.00010	0.000	0.15	EOR	0.00087	0.000	2.77
EAM	0.00049	0.000	0.56	EML	0.00034	0.000	0.65	EOR	0.00012	0.000	0.18	EAN	0.00088	0.000	2.83
EAN	0.00056	0.000	0.65	YRI	0.00040	0.000	0.78	YRI	0.00013	0.000	0.20	EML	0.00088	0.000	2.82
EML	0.00071	0.000	0.82	ASW	0.00044	0.000	0.84	EML	0.00015	0.000	0.23	ASW	0.00093	0.000	2.97

SD, standard deviation of π; K, average number of pairwise differences. ASW, African American; CEU, European; CHB, Han Chinese; CHS, Southern Han Chinese; CLM, Columbian; FIN, Finnish; GBR, British; IBS, Spanish; JPT, Japanese; LWK, Luhya; MXL, Mexican American; PUR, Puerto Rican; TSI, Tuscan; YRI, Yoruba; EAF, Afar; EAM, Amhara; EAN, Anuak; EML, Maale; EOR, Oromo.

Supplementary Figure S2 Population pairwise genetic distances (lower triangle) and *P* values (upper triangle) for *CYP3A4* entire gene haplotypes

	***CYP3A4* entire gene haplotypes**																		
	ASW	**CEU**	**CHB**	**CHS**	**CLM**	**FIN**	**GBR**	**IBS**	**JPT**	**LWK**	**MXL**	**PUR**	**TSI**	**YRI**	**EAF**	**EAM**	**EAN**	**EML**	**EOR**
ASW		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0500	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
CEU	0.3077		<0.0001	<0.0001	<0.0001	0.5100	0.4100	0.6500	<0.0001	<0.0001	<0.0001	<0.0001	0.5900	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
CHB	0.1478	0.0888		0.5400	<0.0100	<0.0001	<0.0001	<0.0500	<0.0100	<0.0001	<0.0100	<0.0500	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
CHS	0.1688	0.0663	-0.0011		<0.0500	<0.0001	<0.0001	0.0800	<0.0001	<0.0001	<0.0001	<0.0500	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
CLM	0.1274	0.0829	0.0304	0.0220		<0.0001	<0.0001	<0.0500	<0.0001	<0.0001	<0.0100	0.1200	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
FIN	0.2945	-0.0013	0.0811	0.0595	0.0666		0.7600	0.6400	<0.0001	<0.0001	<0.0001	<0.0001	0.7300	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
GBR	0.3013	-0.0007	0.0837	0.0619	0.0741	-0.0030		0.5500	<0.0001	<0.0001	<0.0001	<0.0001	0.5300	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
IBS	0.1966	-0.0072	0.0391	0.0230	0.0304	-0.0086	-0.0067		0.0600	<0.0001	<0.0001	0.0600	0.8600	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
JPT	0.1782	0.0817	0.0147	0.0211	0.0456	0.0741	0.0772	0.0382		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
LWK	0.0281	0.3731	0.2210	0.2443	0.1958	0.3595	0.3646	0.2715	0.2511		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
MXL	0.1088	0.1266	0.0243	0.0237	0.0186	0.1167	0.1204	0.0528	0.0550	0.1768		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
PUR	0.1121	0.0744	0.0157	0.0104	0.0045	0.0633	0.0691	0.0254	0.0353	0.1864	0.0198		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
TSI	0.2961	-0.0019	0.0765	0.0556	0.0707	-0.0029	-0.0018	-0.0111	0.0711	0.3621	0.1122	0.0630		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
YRI	0.0077	0.3604	0.2048	0.2283	0.1846	0.3486	0.3538	0.2555	0.2357	0.0110	0.1603	0.1709	0.3503		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
EAF	0.0415	0.2104	0.0832	0.0964	0.0661	0.1977	0.2048	0.1268	0.1060	0.0726	0.0630	0.0557	0.1991	0.0719		<0.0001	0.5500	<0.0001	0.2300
EAM	0.0472	0.3773	0.2188	0.2421	0.1935	0.3655	0.3718	0.2642	0.2480	0.0341	0.1742	0.1824	0.3678	0.0390	0.0574		<0.0001	<0.0001	<0.0001
EAN	0.0353	0.2149	0.0867	0.1008	0.0658	0.2019	0.2089	0.1288	0.1062	0.0671	0.0607	0.0588	0.2036	0.0645	-0.0014	0.0519		<0.0001	0.3500
EML	0.0263	0.3104	0.1556	0.1774	0.1344	0.2981	0.3048	0.2038	0.1816	0.0373	0.1177	0.1254	0.3001	0.0335	0.0278	0.0160	0.0206		<0.0001
EOR	0.0310	0.2205	0.0909	0.1054	0.0690	0.2071	0.2142	0.1322	0.1130	0.0641	0.0648	0.0612	0.2088	0.0591	0.0015	0.0477	0.0006	0.0181	

P values below and above the 5% significance threshold are highlighted in red and blue respectively. Populations: ASW, African American; CEU, European; CHB, Han Chinese; CHS, Southern Han Chinese; CLM, Columbian; FIN, Finnish; GBR, British; IBS, Spanish; JPT, Japanese; LWK, Luhya; MXL, Mexican American; PUR, Puerto Rican; TSI, Tuscan; YRI, Yoruba; EAF, Afar; EAM, Amhara; EAN, Anuak; EML, Maale; EOR, Oromo.

Supplementary Figure S3 Population pairwise genetic distances (lower triangle) and *P* values (upper triangle) for *CYP3A4* exonic haplotypes

	***CYP3A4* exonic haplotypes**																		
	ASW	**CEU**	**CHB**	**CHS**	**CLM**	**FIN**	**GBR**	**IBS**	**JPT**	**LWK**	**MXL**	**PUR**	**TSI**	**YRI**	**EAF**	**EAM**	**EAN**	**EML**	**EOR**
ASW		<0.050	<0.010	0.063	0.162	<0.050	<0.050	0.342	0.117	0.360	<0.050	0.99	<0.010	0.153	0.360	0.505	0.324	<0.050	0.189
CEU	0.032		0.459	0.342	0.676	0.162	0.162	0.991	<0.050	<0.001	0.991	<0.010	0.991	<0.001	0.063	<0.001	<0.001	<0.001	<0.001
CHB	0.048	0.001		0.252	0.180	<0.050	<0.050	0.991	<0.001	<0.001	0.441	<0.010	0.477	<0.001	<0.050	<0.001	<0.001	<0.001	<0.001
CHS	0.021	0.000	0.006		0.649	0.225	0.162	0.991	0.180	<0.001	0.559	<0.050	0.604	<0.001	0.180	<0.001	<0.001	<0.001	<0.001
CLM	0.009	-0.002	0.010	-0.002		0.775	0.477	0.991	0.198	<0.010	0.604	0.144	0.856	<0.010	0.505	0.108	<0.050	<0.001	<0.050
FIN	0.011	0.004	0.013	0.001	-0.003		0.459	0.991	0.189	<0.050	0.225	<0.010	0.198	<0.001	0.378	<0.050	<0.001	<0.001	<0.010
GBR	0.014	0.004	0.014	0.001	-0.002	-0.001		0.991	0.126	<0.001	0.288	<0.050	0.207	<0.001	0.333	<0.050	<0.001	<0.001	<0.001
IBS	0.005	-0.018	0.000	-0.014	-0.015	-0.010	-0.010		0.613	0.252	0.991	0.396	0.991	0.072	0.703	0.387	0.144	0.135	0.198
JPT	0.008	0.015	0.024	0.006	0.004	0.003	0.005	-0.004		<0.010	0.090	0.081	<0.050	<0.001	0.225	0.081	<0.001	<0.001	<0.050
LWK	-0.001	0.040	0.051	0.031	0.020	0.021	0.024	0.013	0.015		<0.001	0.505	<0.001	0.099	0.081	0.279	0.333	<0.050	0.180
MXL	0.023	-0.003	0.003	-0.001	-0.002	0.003	0.003	-0.017	0.010	0.033		<0.050	0.991	<0.001	0.135	<0.010	<0.001	<0.001	<0.001
PUR	-0.008	0.039	0.057	0.026	0.013	0.015	0.016	0.008	0.010	-0.003	0.030		<0.050	0.099	0.261	0.378	0.450	0.126	0.144
TSI	0.022	-0.002	0.003	-0.001	-0.004	0.002	0.003	-0.016	0.011	0.033	-0.003	0.029		<0.001	0.189	<0.001	<0.001	<0.001	<0.001
YRI	0.010	0.072	0.086	0.064	0.045	0.048	0.052	0.033	0.038	0.006	0.060	0.008	0.066		<0.001	0.06	0.505	0.568	0.108
EAF	0.001	0.010	0.020	0.004	-0.002	0.000	0.001	-0.008	0.002	0.011	0.006	0.003	0.004	0.033		0.387	<0.050	<0.010	<0.050
EAM	-0.002	0.028	0.040	0.019	0.011	0.010	0.013	0.003	0.005	0.002	0.020	-0.002	0.023	0.012	0.002		0.126	<0.050	0.766
EAN	0.000	0.061	0.077	0.050	0.033	0.036	0.040	0.024	0.027	-0.001	0.050	-0.002	0.053	-0.002	0.021	0.006		0.297	0.126
EML	0.016	0.084	0.101	0.075	0.053	0.057	0.062	0.038	0.045	0.009	0.071	0.011	0.078	-0.002	0.040	0.018	0.000		0.072
EOR	0.004	0.043	0.056	0.034	0.023	0.021	0.025	0.012	0.014	0.003	0.034	0.003	0.038	0.008	0.010	-0.004	0.006	0.011	

P values below and above the 5% significance threshold are highlighted in red and blue respectively. Populations: ASW, African American; CEU, European; CHB, Han Chinese; CHS, Southern Han Chinese; CLM, Columbian; FIN, Finnish; GBR, British; IBS, Spanish; JPT, Japanese; LWK, Luhya; MXL, Mexican American; PUR, Puerto Rican; TSI, Tuscan; YRI, Yoruba; EAF, Afar; EAM, Amhara; EAN, Anuak; EML, Maale; EOR, Oromo.

Supplementary Figure S4 Population pairwise genetic distances (lower triangle) and *P* values (upper triangle) for *CYP3A4* 3' UTR haplotypes

	CYP3A4 3' UTR haplotypes																		
	ASW	**CEPH**	**CHB**	**CHS**	**CLM**	**FIN**	**GBR**	**IBS**	**JPT**	**LWK**	**MXL**	**PUR**	**TSI**	**YRI**	**EAF**	**EAN**	**EAM**	**EML**	**EOR**
ASW		0.1500	<0.0001	<0.0001	0.0600	0.0600	<0.0500	0.6000	<0.0001	0.0700	<0.0500	0.2700	0.0800	0.4400	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
CEPH	0.0059		<0.0001	<0.0001	<0.0100	0.4300	0.1500	0.9900	<0.0001	0.1400	0.1900	0.4000	0.6200	0.1500	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
CHB	0.0616	0.0346		0.4800	0.5900	<0.0001	<0.0500	<0.0500	0.6700	<0.0001	<0.0500	<0.0500	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
CHS	0.0749	0.0436	0.0002		0.1400	<0.0001	<0.0100	<0.0100	0.2900	<0.0001	<0.0100	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
CLM	0.0274	0.0145	0.0011	0.0100		0.1400	0.5200	0.2300	0.6300	0.1200	0.5000	0.3600	0.1400	0.0900	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
FIN	0.0117	-0.0019	0.0254	0.0340	0.0067		0.5400	0.7600	<0.0500	0.6400	0.3500	0.4600	0.9900	0.2600	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
GBR	0.0194	0.0037	0.0110	0.0186	-0.0020	-0.0016		0.2900	0.0900	0.2300	0.9400	0.6700	0.4400	0.2300	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
IBS	-0.0090	-0.0197	0.1010	0.1722	0.0254	-0.0162	-0.0026		<0.0500	0.5100	0.2300	0.6200	0.9900	0.5200	0.0600	<0.0001	<0.0001	<0.0001	<0.0500
JPT	0.0483	0.0261	-0.0031	0.0067	-0.0015	0.0181	0.0058	0.0587		<0.0100	0.1900	0.0900	<0.0500	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
LWK	0.0137	0.0039	0.0328	0.0422	0.0109	-0.0027	0.0018	-0.0100	0.0248		0.2300	0.2200	0.3600	0.2100	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
MXL	0.0203	0.0062	0.0120	0.0223	-0.0023	0.0005	-0.0053	0.0030	0.0058	0.0031		0.5200	0.2900	0.2600	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
PUR	0.0040	-0.0009	0.0212	0.0337	0.0004	-0.0018	-0.0035	-0.0099	0.0121	0.0035	-0.0016		0.3900	0.7100	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
TSI	0.0112	-0.0028	0.0260	0.0343	0.0079	-0.0051	-0.0007	-0.0174	0.0189	-0.0015	0.0016	-0.0017		0.2900	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
YRI	-0.0009	0.0025	0.0289	0.0378	0.0073	0.0007	0.0029	-0.0098	0.0210	0.0022	0.0037	-0.0044	0.0011		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
EAF	0.0824	0.1014	0.1918	0.2047	0.1426	0.1257	0.1411	0.0607	0.1745	0.1330	0.1320	0.1093	0.1243	0.1159		<0.0001	0.45	<0.0001	0.7400
EAN	0.2312	0.2717	0.3719	0.3850	0.3039	0.3024	0.3180	0.1881	0.3515	0.3096	0.2961	0.2660	0.3032	0.2872	0.0624		<0.0001	0.0500	<0.0001
EAM	0.0875	0.1096	0.1984	0.2106	0.1487	0.1344	0.1496	0.0640	0.1815	0.1415	0.1389	0.1157	0.1334	0.1241	-0.0025	0.0478		<0.0100	0.9500
EML	0.1786	0.2146	0.3100	0.3223	0.2472	0.2441	0.2592	0.1387	0.2909	0.2518	0.2395	0.2112	0.2445	0.2305	0.0414	0.0046	0.0247		<0.0100
EOR	0.0835	0.1046	0.1957	0.2084	0.1453	0.1296	0.1452	0.0604	0.1783	0.1369	0.1350	0.1116	0.1284	0.1194	-0.0043	0.0537	-0.0057	0.0291	

P values below and above the 5% significance threshold are highlighted in red and blue respectively. Populations: ASW, African American; CEU, European; CHB, Han Chinese; CHS, Southern Han Chinese; CLM, Columbian; FIN, Finnish; GBR, British; IBS, Spanish; JPT, Japanese; LWK, Luhya; MXL, Mexican American; PUR, Puerto Rican; TSI, Tuscan; YRI, Yoruba; EAF, Afar; EAM, Amhara; EAN, Anuak; EML, Maale; EOR, Oromo.

Supplementary Figure S5 Population pairwise genetic distances (lower triangle) and *P* values (upper triangle) for *CYP3A4* 5' regulatory region haplotypes

	CYP3A4 5' regulatory region haplotypes																		
	ASW	**CEPH**	**CHB**	**CHS**	**CLM**	**FIN**	**GBR**	**IBS**	**JPT**	**LWK**	**MXL**	**PUR**	**TSI**	**YRI**	**EAF**	**EAN**	**EAM**	**EML**	**EOR**
ASW		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0500	<0.0001	<0.0001	<0.0001	<0.0100	<0.0001
CEPH	0.5079		0.1300	0.1200	<0.0001	0.5400	0.3200	0.9900	0.1400	<0.0001	0.1300	<0.0001	0.3100	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
CHB	0.5569	0.0128		0.9900	<0.0001	<0.0500	<0.0001	0.9900	0.4700	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
CHS	0.5616	0.0132	0.0000		<0.0001	<0.0001	<0.0001	0.9900	0.4100	<0.0001	<0.0001	<0.0001	<0.0100	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
CLM	0.3333	0.0644	0.1218	0.1240		<0.0500	<0.0500	0.0800	<0.0001	<0.0001	<0.0500	0.0800	<0.0500	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
FIN	0.4887	-0.0010	0.0278	0.0284	0.0371		0.7900	0.5600	<0.0500	<0.0001	0.2900	<0.0001	0.9900	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
GBR	0.4707	0.0031	0.0358	0.0365	0.0265	-0.0048		0.5900	<0.0500	<0.0001	0.4400	<0.0001	0.7900	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
IBS	0.3645	-0.0108	0.0000	0.0000	0.0440	-0.0016	0.0028		0.9900	<0.0001	0.3900	<0.0500	0.6300	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
JPT	0.5349	0.0046	0.0005	0.0007	0.1015	0.0179	0.0249	-0.0177		<0.0001	<0.0100	<0.0001	<0.0500	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
LWK	0.0709	0.6638	0.6980	0.7013	0.5286	0.6498	0.6366	0.5652	0.6829		<0.0001	<0.0001	<0.0001	<0.0500	<0.0001	0.1200	<0.0001	<0.0001	<0.0001
MXL	0.4135	0.0104	0.0437	0.0447	0.0161	0.0003	-0.0019	0.0038	0.0308	0.5954		<0.0001	0.3700	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
PUR	0.2208	0.1646	0.2311	0.2347	0.0203	0.1300	0.1124	0.1040	0.2066	0.4248	0.0837		<0.0001	<0.0001	<0.0001	<0.0001	<0.0500	<0.0001	<0.0001
TSI	0.4903	-0.0003	0.0262	0.0267	0.0353	-0.0050	-0.0046	-0.0021	0.0171	0.6510	-0.0008	0.1278		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
YRI	0.0144	0.5687	0.6071	0.6109	0.4228	0.5542	0.5396	0.4537	0.5897	0.0156	0.4924	0.3180	0.5561		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
EAF	0.0669	0.3141	0.3638	0.3679	0.1531	0.2909	0.2737	0.2174	0.3435	0.2234	0.2316	0.0653	0.2910	0.1405		<0.0001	0.4700	0.1200	0.8900
EAN	0.1155	0.7382	0.7736	0.7766	0.5970	0.7223	0.7087	0.6363	0.7582	0.0057	0.6669	0.4876	0.7225	0.0455	0.2672		<0.0001	<0.0001	<0.0001
EAM	0.0910	0.2755	0.3257	0.3296	0.1192	0.2513	0.2343	0.1882	0.3057	0.2635	0.1951	0.0401	0.2512	0.1735	-0.0020	0.3099		<0.0500	0.2600
EML	0.0300	0.4139	0.4641	0.4685	0.2401	0.3916	0.3734	0.2995	0.4431	0.1435	0.3260	0.1331	0.3925	0.0802	0.0100	0.1798	0.0256		0.1400
EOR	0.0639	0.3362	0.3880	0.3923	0.1679	0.3122	0.2942	0.2334	0.3669	0.2160	0.2502	0.0749	0.3122	0.1350	-0.0057	0.2597	0.0018	0.0067	

P values below and above the 5% significance threshold are highlighted in red and blue respectively. Populations: ASW, African American; CEU, European; CHB, Han Chinese; CHS, Southern Han Chinese; CLM, Columbian; FIN, Finnish; GBR, British; IBS, Spanish; JPT, Japanese; LWK, Luhya; MXL, Mexican American; PUR, Puerto Rican; TSI, Tuscan; YRI, Yoruba; EAF, Afar; EAM, Amhara; EAN, Anuak; EML, Maale; EOR, Oromo.

Table S5. Results of neutrality tests performed on *CYP3A4* variation in populations of Ethiopia and The 1000 Genomes Project

Population	n	Tajimas D	Fu and Li's Test	
		D (P value)	D* (P value)	F* (P value)
ASW	122	-1.01 (>0.10)	-1.70 (>0.10)	-1.70 (>0.10)
CEU	174	**-1.83 (<0.05)**	-1.12 (>0.10)	-1.66 (>0.10)
CHB	194	-0.90 (>0.10)	**-3.74 (<0.02)**	**-3.23 (<0.02)**
CHS	200	-0.51 (>0.10)	-1.44 (>0.10)	-1.32 (>0.10)
CLM	120	-1.56 (>0.05)	**-3.61 (<0.02)**	**-3.38 (<0.02)**
FIN	186	**-1.80 (<0.05)**	-1.45 (>0.10)	-1.90 (>0.10)
GBR	178	**-1.92 (<0.05)**	**-3.35 (<0.02)**	**-3.36 (<0.02)**
IBS	28	-1.06 (>0.10)	-0.58 (>0.10)	-0.83 (>0.10)
JPT	178	-0.73 (>0.10)	-0.19 (>0.10)	-0.46 (>0.10)
LWK	194	-1.31 (>0.10)	-0.77 (>0.10)	-1.96 (>0.10)
MXL	132	-1.15 (>0.10)	-2.22 (>0.05)	-2.17 (>0.05)
PUR	110	-1.48 (>0.10)	**-3.10 (<0.05)**	**-2.95 (<0.05)**
TSI	196	**-1.95 (<0.05)**	**-3.54 (<0.02)**	**-3.51 (<0.02)**
YRI	176	-1.19 (>0.10)	-0.57 (>0.10)	-0.99 (>0.10)
EAF	152	-0.50 (>0.10)	-0.54 (>0.10)	-0.63 (>0.10)
EAM	152	-1.00 (>0.10)	-1.06 (>0.10)	-1.24 (>0.10)
EAN	152	-0.75 (>0.10)	-0.15 (>0.10)	-0.48 (>0.10)
EML	150	-0.87 (>0.10)	-1.78 (>0.10)	-1.68 (>0.10)
EOR	146	-1.08 (>0.10)	**-2.39 (<0.05)**	-2.22 (>0.05)

All *CYP3A4* variants, including singletons, were included; n represents the number of chromosomes.

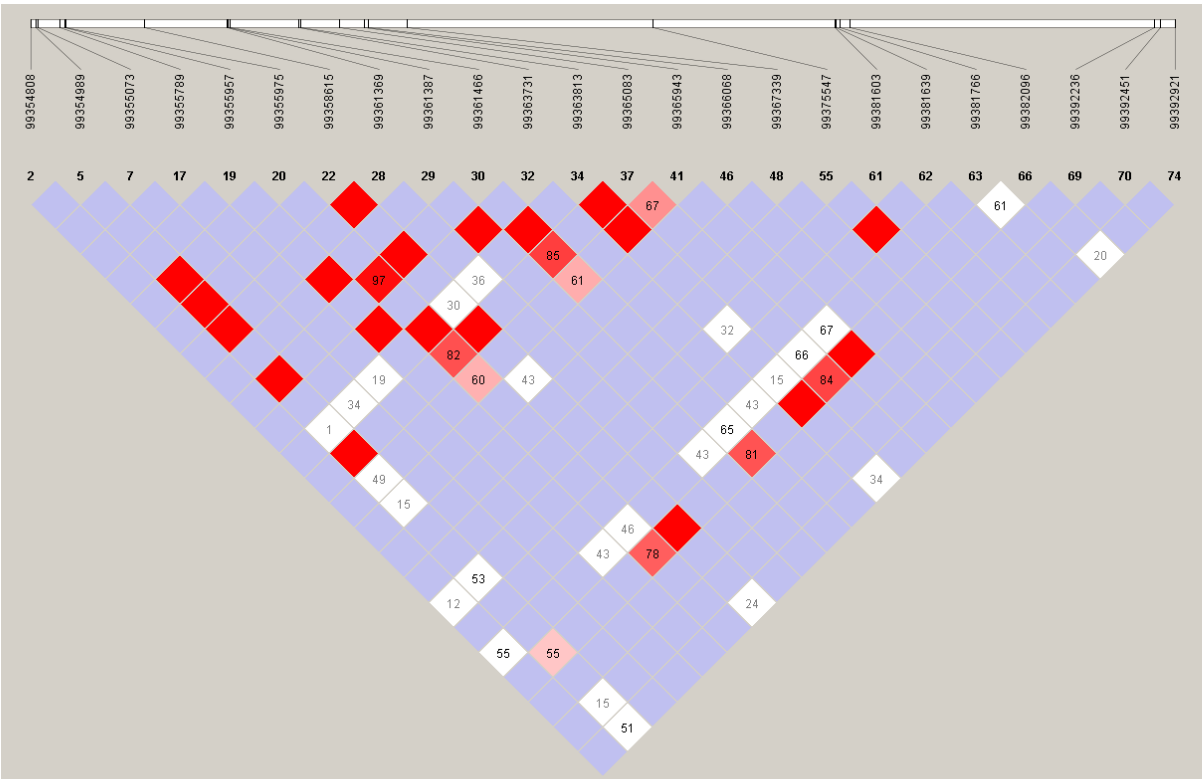
Supplementary Material S4

LD plots are provided for each individual population and for the combined world dataset. The strength of LD is based on D’ and LOD (logarithm of the odds) scores and increases with colour from white to blue to red as depicted in Key 1. The first row of numbers along the top of the plot denotes the chromosome position in Human Reference Assembly 36.2 of each variant (see Tables 1 and 2). The space between loci is indicated by the white bar at the top of the plot. Numbers within squares are D’ values (multiplied 100-fold). Empty squares indicate D’ of 1. Singletons and low frequency (<0.001) variants were excluded from analysis

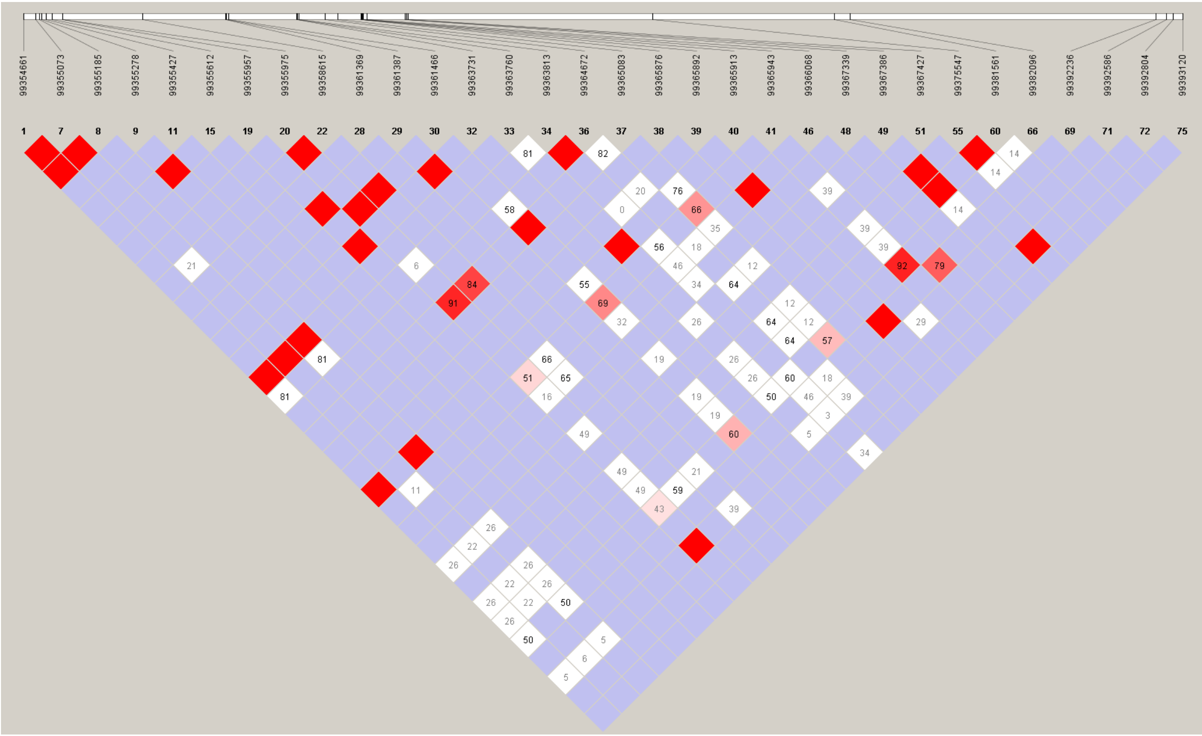
Key 1

	D' <1	D' = 1
LOD ≤ 2	White	Blue
LOD ≥ 2	Shade of red/pink	Bright Red

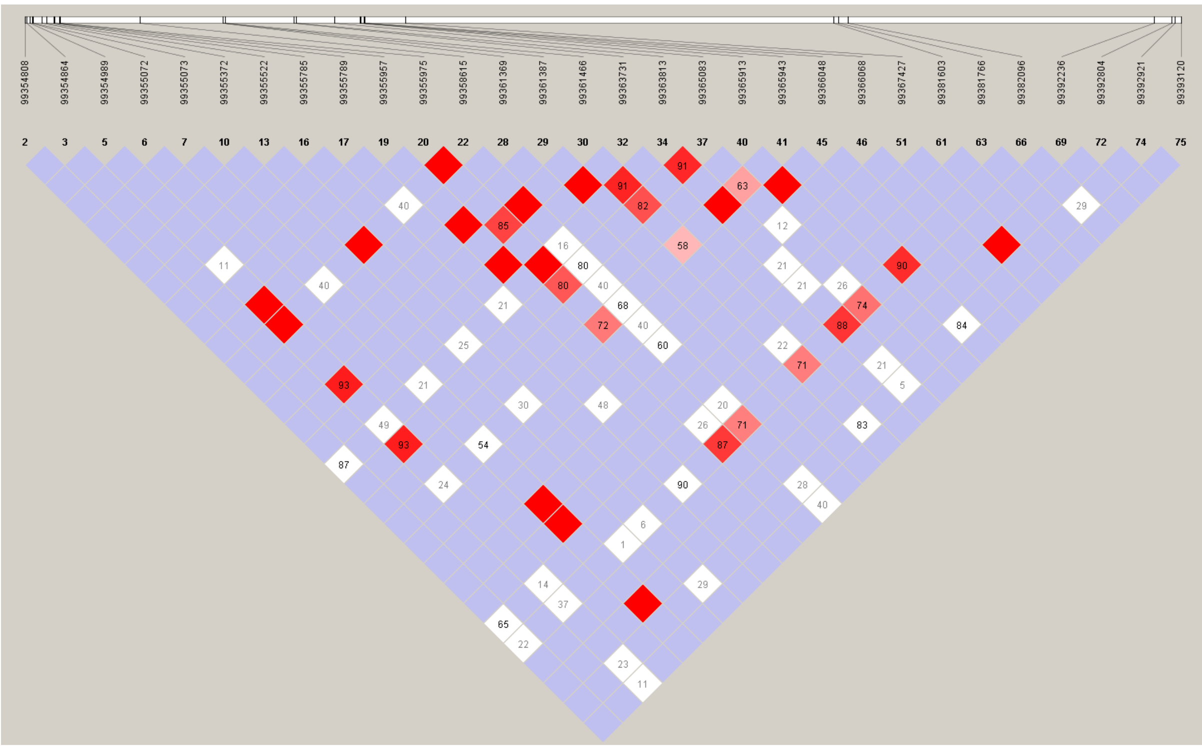
Afar



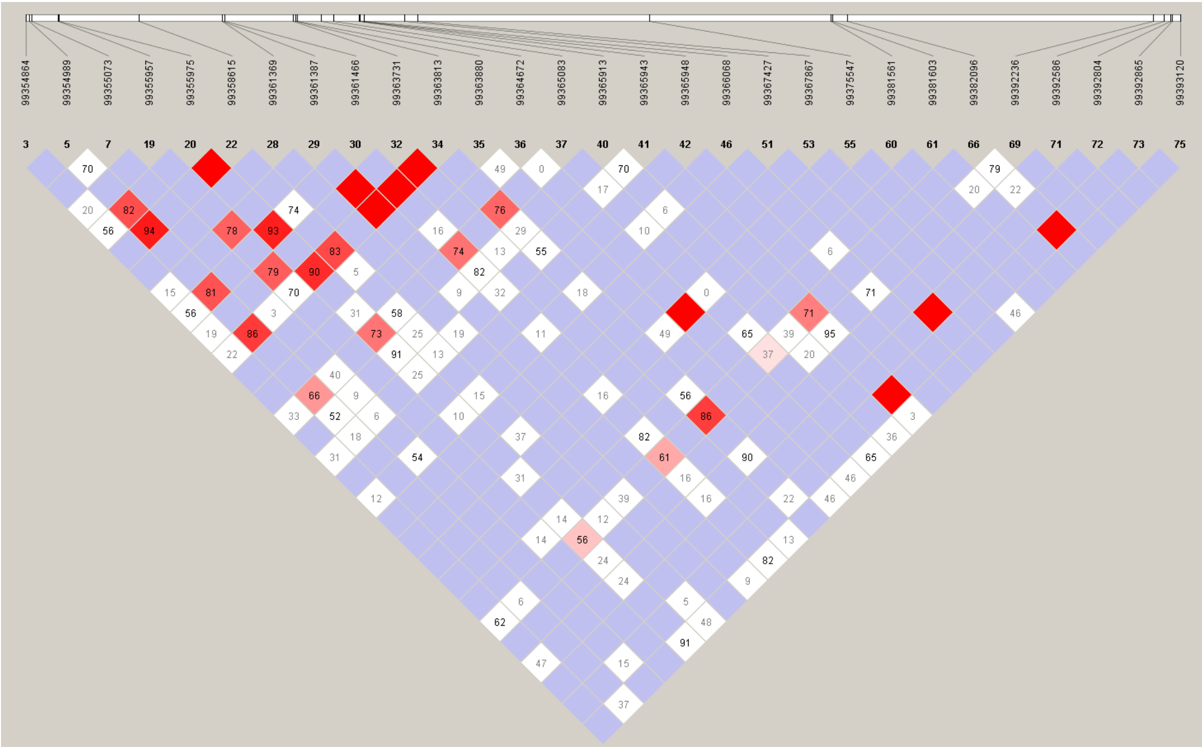
African American



Amhara

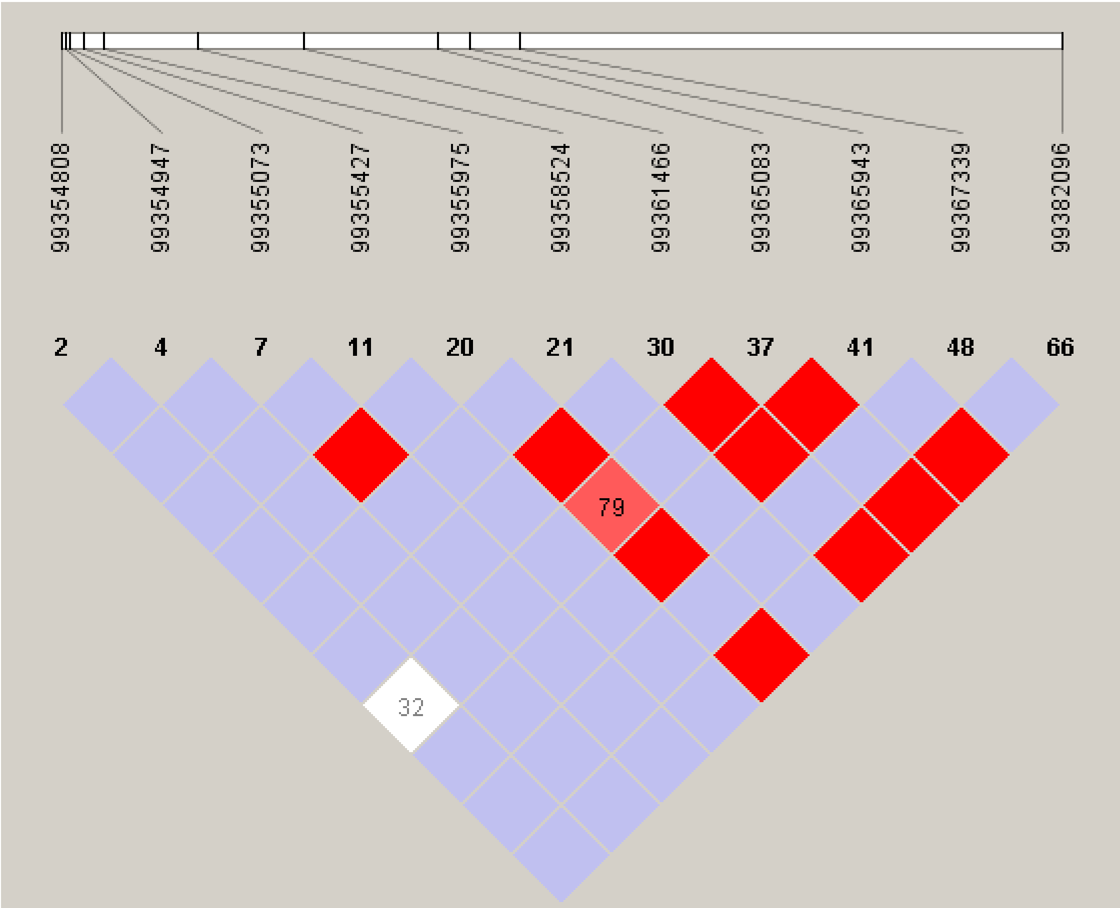


Anuak



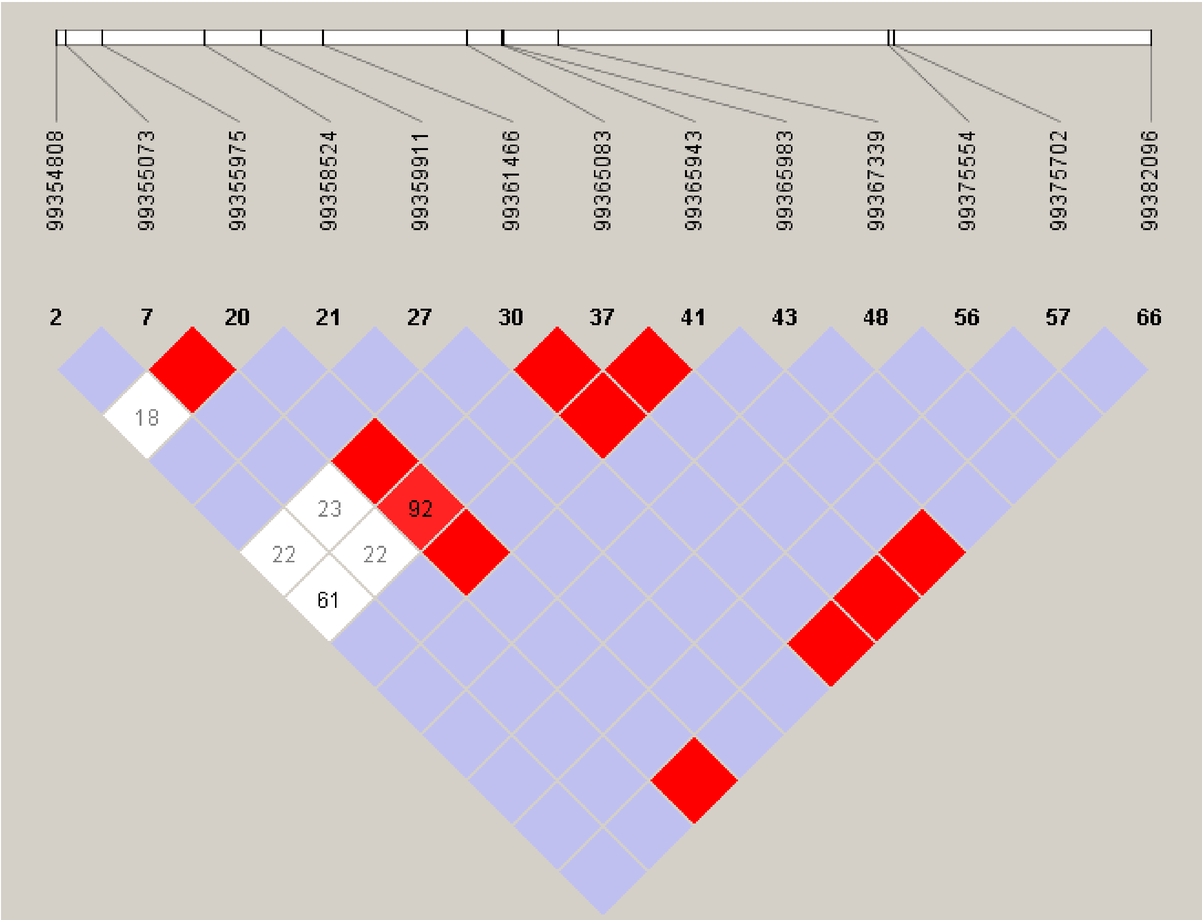
British

CEU European

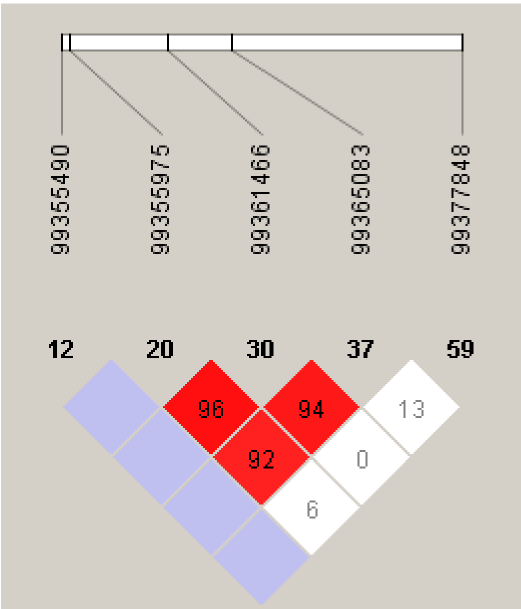


Columbian

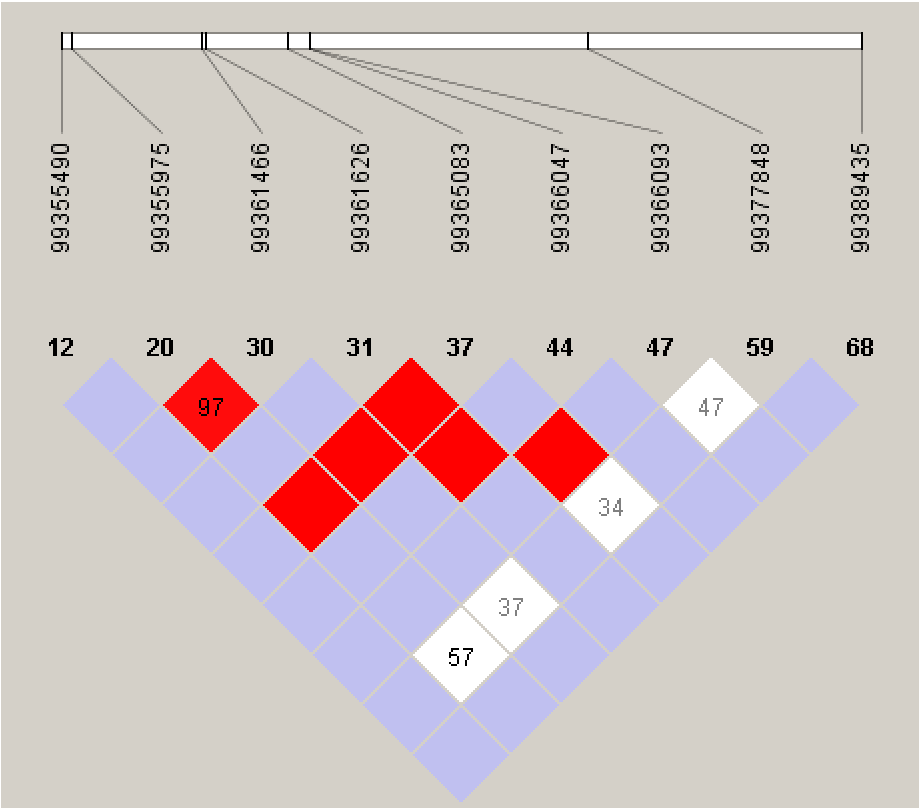
Finnish



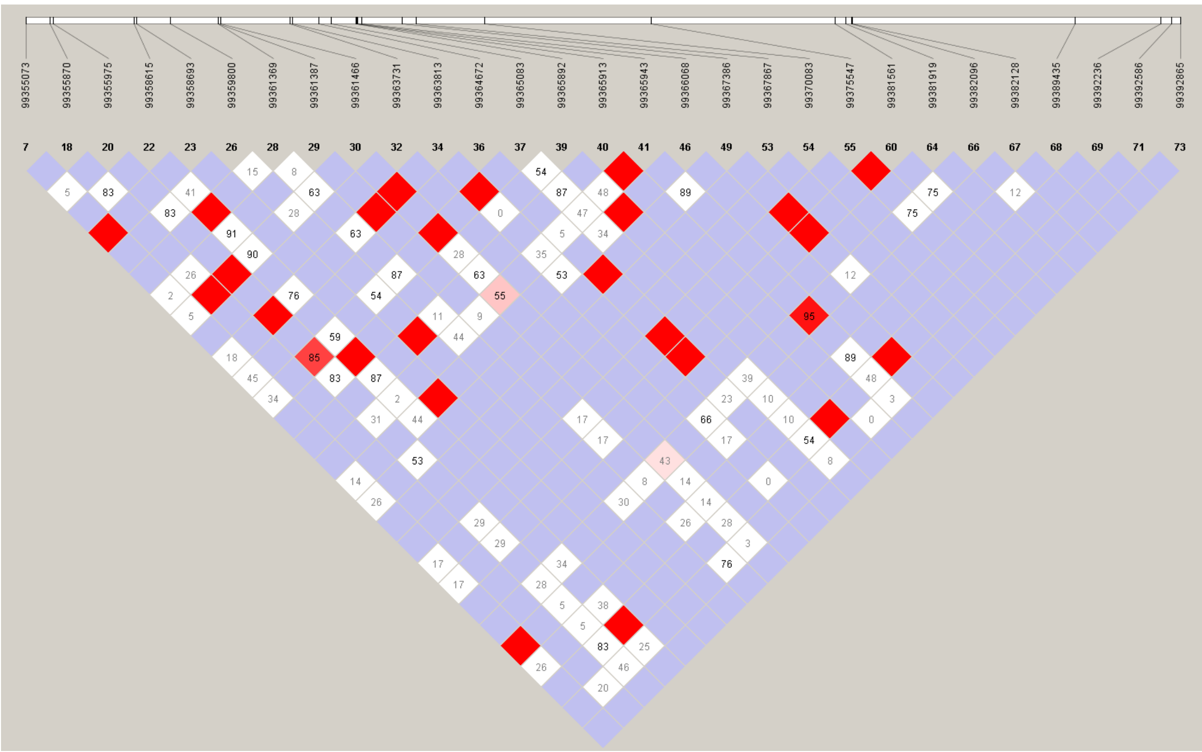
Han Chinese



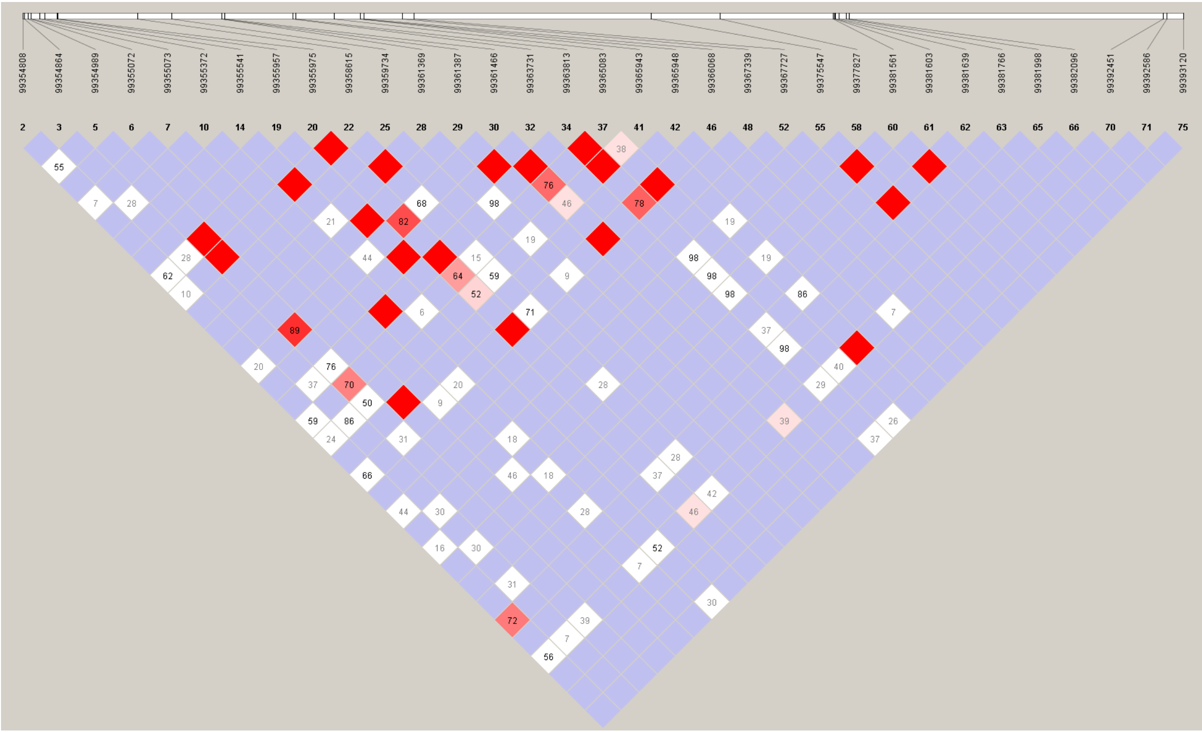
Japanese



Luhya

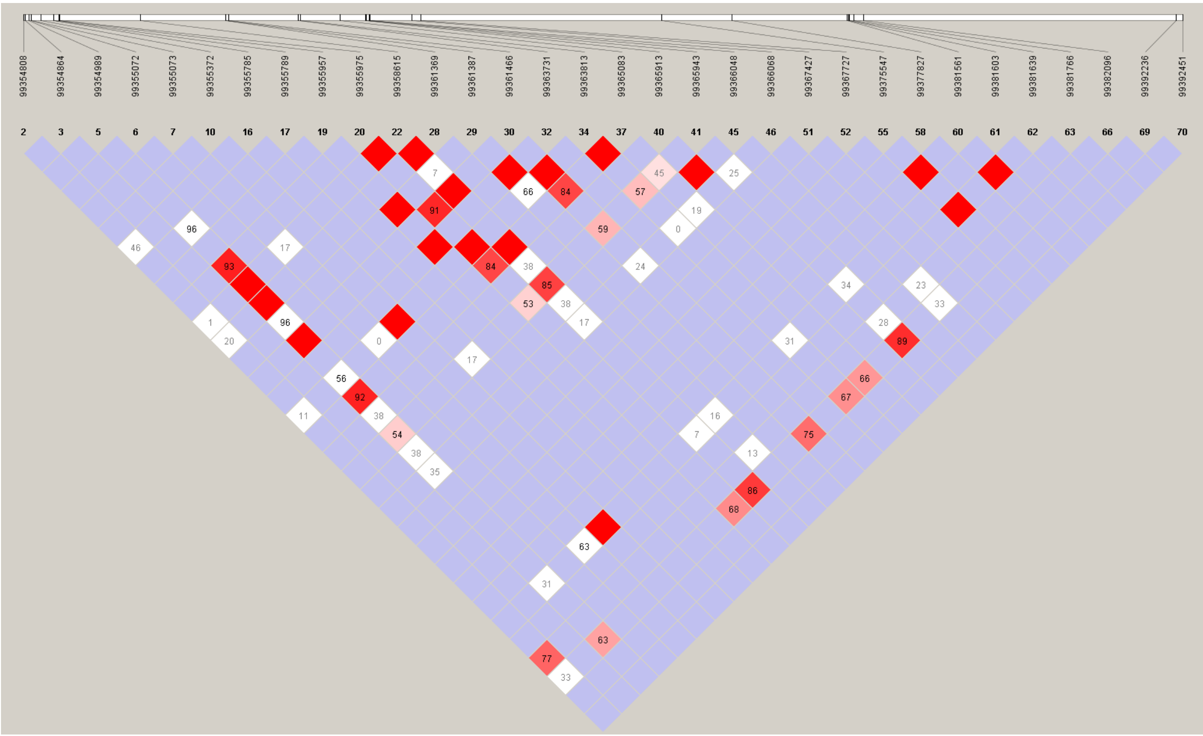


Maale

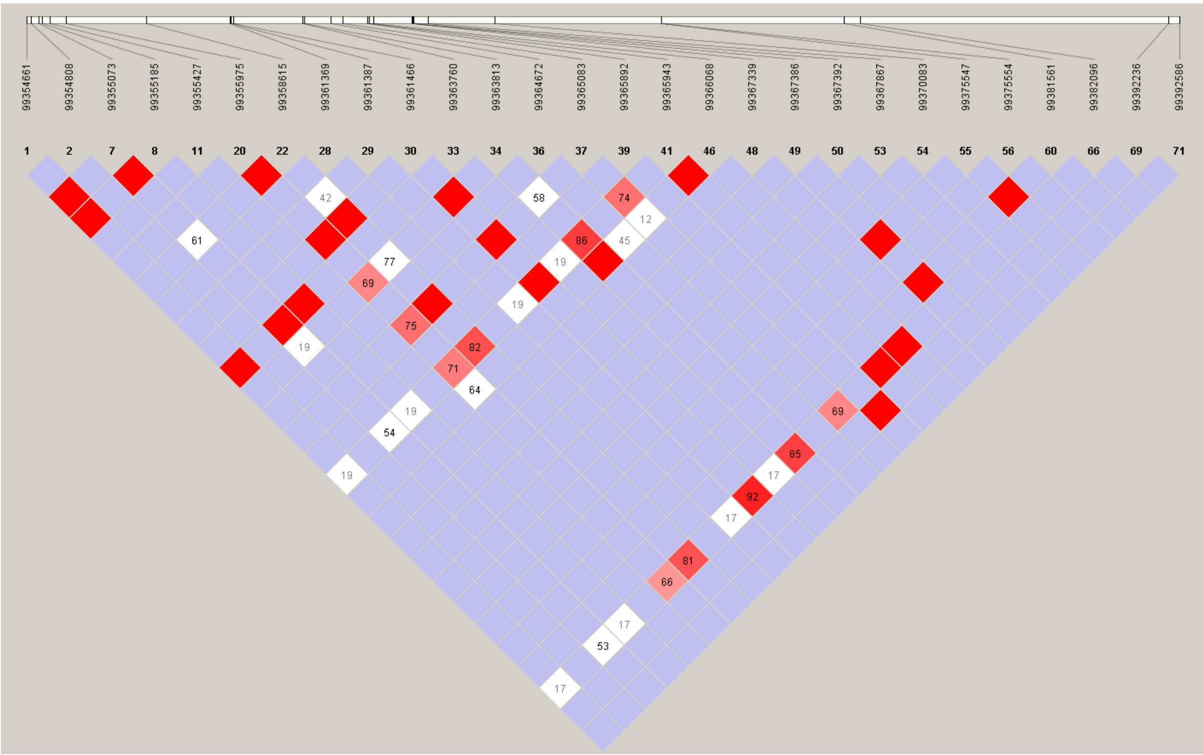


Mexican American

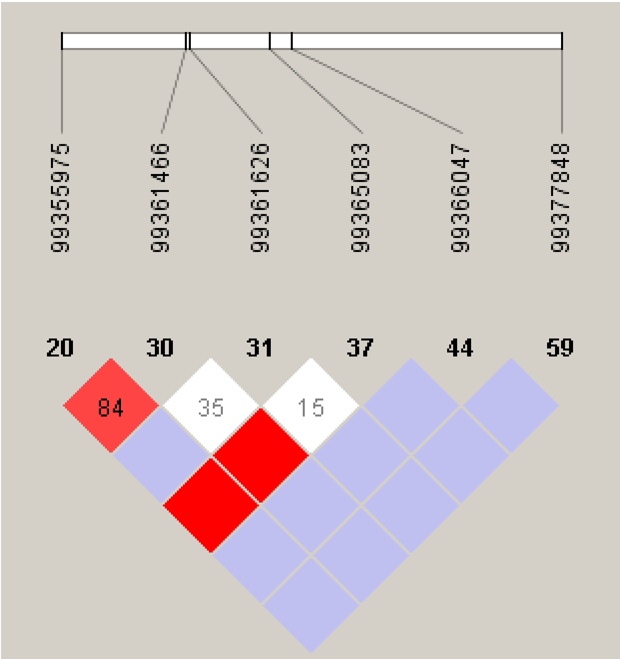
Oromo



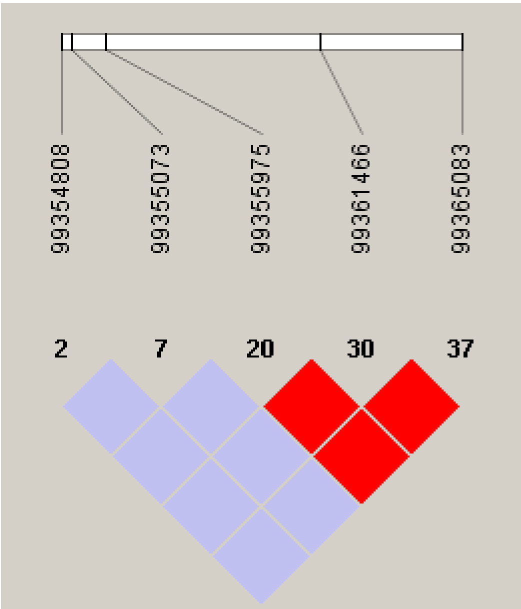
Puerto Rican



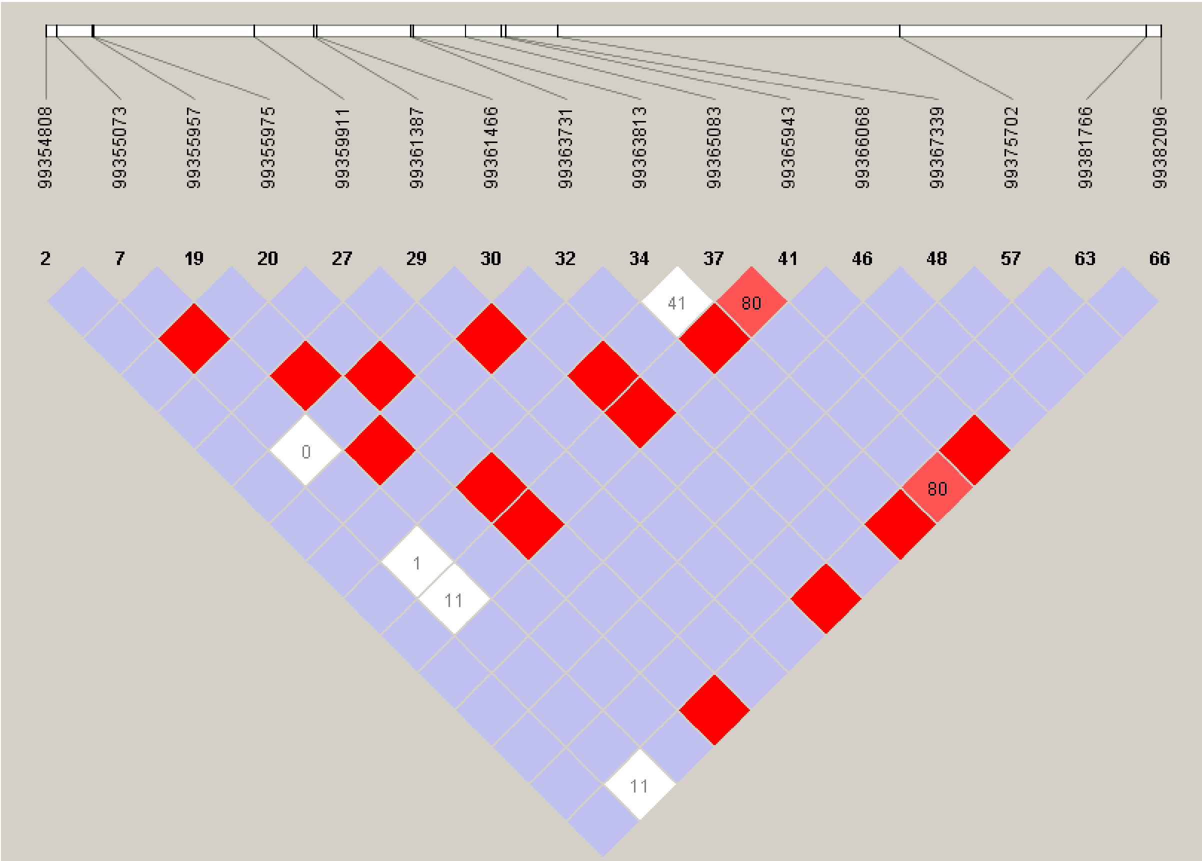
Southern Han Chinese



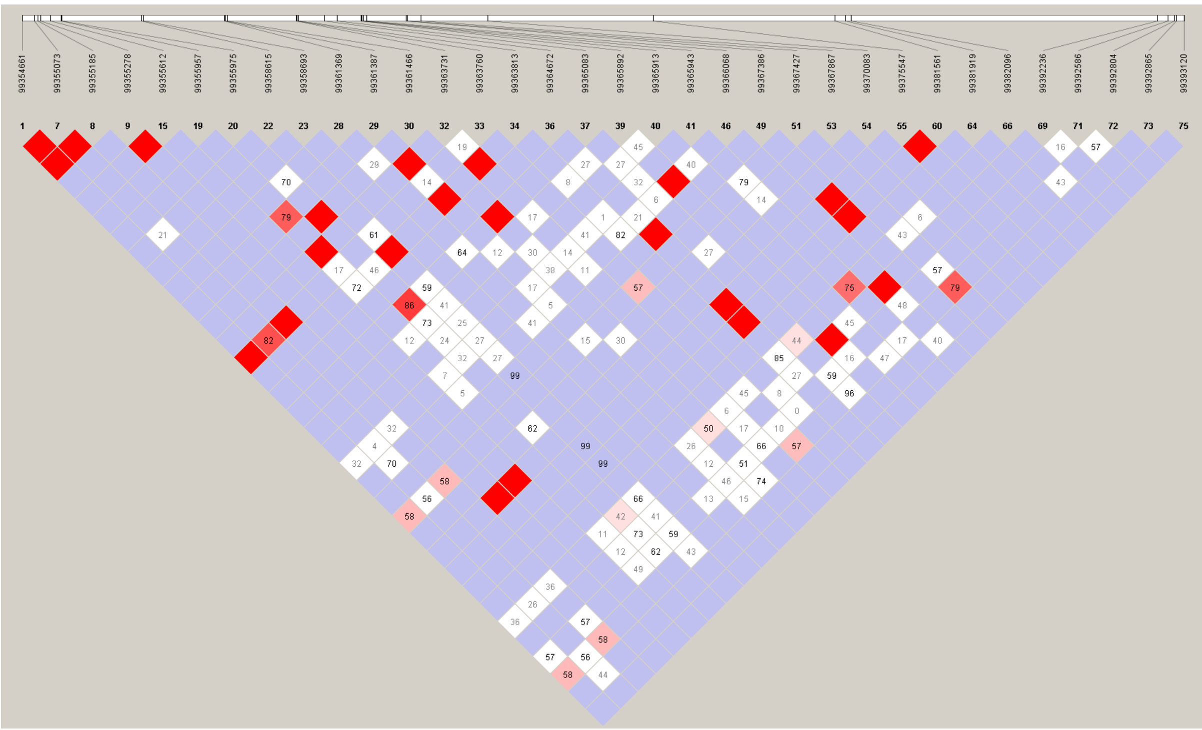
Spanish



Tuscan



Yoruba



World dataset