HHS Public Access

Author manuscript

Pharmacogenet Genomics. Author manuscript; available in PMC 2017 May 01.

Published in final edited form as:

Pharmacogenet Genomics. 2016 May ; 26(5): 197–207. doi:10.1097/FPC.0000000000000206.

Regulatory effects of genomic translocations at the human carboxylesterase-1 (*CES1*) gene locus

Jonathan Christian Sanford[1], **Xinwen Wang**[2], **Jian Shi**[2], **Elizabeth Stofko Barrie**[1], **Danxin Wang**[1], **Hao-Jie Zhu**[2], and **Wolfgang Sadee**[1]

[1]Center for Pharmacogenomics, College of Medicine, The Ohio State University, Columbus, OH, USA

[2]Department of Clinical Pharmacy, University of Michigan, Ann Arbor, MI, USA

Abstract

Objectives—*CES1* encodes carboxylesterase-1, an important drug metabolizing enzyme with high expression in the liver. Previous studies have demonstrated a genomic translocation of the 5′ region from the poorly expressed pseudogene *CES1P1,* to *CES1,* yielding structural variant *CES1VAR*. The purpose of this study was to characterize this translocation and its effect on CES1 expression in human liver.

Methods—Experiments were performed in human liver tissues and cell culture (HepG2). The promoter and exon-1 of *CES1* were sequenced by Sanger and Ion Torrent sequencing to identify gene translocations. The effect of *CES1* 5′UTRs on mRNA and protein expression were assessed by quantitative real-time-PCR, allelic ratio mRNA analysis by primer extension (SNaPshot), quantitative targeted proteomics, and luciferase reporter gene assays.

Results—Sequencing of *CES1* identified two translocations. First, *CES1VAR* (17% minor allele frequency, MAF) comprising the 5′UTR, exon-1, and part of intron-1. A second shorter translocation, *CES1SVAR,* was observed excluding exon-1 and intron-1 regions (<0.01% MAF). *CES1VAR* associates with 2.6-fold decreased CES1 mRNA and ~1.35-fold lower allelic mRNA. Luciferase reporter constructs demonstrate *CES1VAR* decreases luciferase activity 1.5-fold, while *CES1SVAR* slightly increases activity. *CES1VAR* was not associated with CES1 protein expression or metabolism of the CES1 substrates enalapril, clopidogrel or methylphenidate in liver.

Conclusions—The frequent translocation variant *CES1VAR* reduces mRNA expression of CES1 in liver by ~30%, but protein expression and metabolizing activity in liver were not detectably altered – possibly due to variable CES1 expression masking small allelic effects. Whether drug therapies are affected by *CES1VAR* will require further in vivo studies.

Correspondence: Wolfgang Sadee, 5078 Graves Hall, the Ohio State University, 333 W. 10th Avenue, Columbus, OH 43210, wolfgang.sadee@osumc.edu, Hao-Jie Zhu, Department of Clinical Pharmacy, University of Michigan College of Pharmacy, 428 Church Street, Room 3567 CCL, Ann Arbor, MI 48109-1065, Tel: 734-763-8449, hjzhu@med.umich.edu.

*Jonathan Christian Sanford and Xinwen Wang are the co-first authors of this article

Conflicts of Interest: None declared

Keywords

CES1; carboxylesterase; translocation; mRNA expression; protein expression; pharmacogenetics; clopidogrel; enalapril; allelic expression imbalance; regulatory

Introduction

A member of the carboxylesterase family, carboxylesterase-1 (CES1) plays a significant role in the metabolism of drugs with ester or amide bonds in the human liver. Common drugs metabolized by CES1 include the anti-platelet prodrug clopidogrel [1], antivirals (oseltamivir) [2, 3], ADHD medications (methylphenidate) [4], chemotherapeutic agents (irinotecan) [5], ACE inhibitors (imidapril, enalapril, trandolapril, ramipril) [6–8] and others. *CES1* is closely related to the pseudogene *CES1P1,* which constitutes less than 2% of total *CES1* hepatic gene expression [9]. Recent studies have shown that polymorphisms within *CES1* can have significant effects on enzyme function and drug efficacy [10, 11]. In the case of clopidogrel, CES1 inactivates 85% of the drug, acting at three levels: the prodrug itself, an intermediate metabolite, and the active metabolite [1]. The relatively rare coding variant rs71647871 (G143E, minor allele frequency (MAF) = 0.01–0.04) has been shown to ablate CES1 metabolic activity [10] and has been proposed as a biomarker to predict clopidogrel efficacy [11] as decreased CES1 metabolic activity leads to increased clopidogrel efficacy. Numerous additional variants have been associated with drug responses or changes in gene expression indicating *CES1* has potential to impact a wide array of substrates (Table 1) [8, 11–14]. While regulatory variants affecting transcription and RNA biology have been identified for a number of drug metabolizing enzymes, no systematic study has been carried out with *CES1* [15, 16].

Previous studies have characterized the structure of the *CES1* gene family locus which consists of *CES1* and *CES1P1* in a tail-to-tail configuration (Figure 1) [9, 17]. Showing high sequence identity with *CES1*, the pseudogene *CES1P1* expresses at low levels either a non-functional truncated mRNA or the protein coding *CES1P1VAR* mRNA. Sequence differences between *CES1* and *CES1P1* exist primarily in the 5′UTR, exon 1, and intron 1 regions, as well as at a premature stop codon within the exon 3 of *CES1P1* (Figure 2, Table 2). *CES1* consists of at least two alleles [9, 18], the wild-type and a variant allele carrying a translocation from *CES1P1* (5′UTR, exon 1 and intron 1 regions) which replaces *CES1* sequence. Referred to as *CES1 variant (CES1VAR)* (Figure 1, Figure 2, Table 2), this isoform consists of 11 highly linked singe nucleotide polymorphisms (SNPs) whose sequence matches the *CES1P1VAR* gene locus. Since the non-synonymous variants in the protein coding portion of exon 1 included in *CES1VAR* do not occur at *CES1* catalytic sites, we hypothesized that *CES1VAR* expression could be affected by the acquired *CES1P1VAR* sequence, now driven by the expression of the functional *CES1* promoter.

Molecular genetic studies testing variants within *CES1* for function or association with changes in gene and protein expression are limited [8, 18]. Studies testing major *CES1* variants (such as *CES1VAR*), have not typically found significant associations with drug response or toxicity [19, 20]. In one case *CES1P1VAR* expression has been related to the

activation of irinotecan to CPT-11 [18]. Further, an additional sequence variant similar to *CES1VAR* has also been identified but is limited to *CES1P1* SNPs in the protein coding region of *CES1* exon 1 [18]. Adding complexity to this region is the inclusion of multiple transcription start sites along the annotated 5′UTR [17]. It is unclear whether and to what extent *CES1* translocation variation affects mRNA isoforms or expression.

The goal of this study is to better understand variation at the 5′ region of the *CES1* gene locus. Further, we aim to define how this 5′ variation affects mRNA expression, mRNA isoforms, and protein expression, to determine whether *CES1* 5′UTR sequences such as *CES1VAR* can serve as biomarkers when CES1 is a primary metabolic enzyme determining drug response.

Materials and Methods

Tissue samples

A total of 227 human liver samples were obtained from the Cooperative Human Tissue Network (Midwest and Western Division) and XenoTech LLC (Lenexa, KS), under approval of the Ohio State University (n=125, primarily Caucasians and 15% with African ancestry, 45% male and 55% female) and the University of Michigan (n=102, 92 Caucasians, 6 African-Americans, 2 Hispanics, and 2 classified as ‘others’, 45% male and 55% female) Institutional Review Boards. An additional fifty liver samples were obtained from other sources that included the Medical College of Virginia (Richmond, VA, USA), Medical College of Wisconsin (Milwaukee, WI, USA), Indiana University School of Medicine (Indianapolis, IN, USA), and University of Pittsburgh (Pittsburgh, PA, USA). All livers were used under protocols approved by committees for the conduct of human research at the Ohio State University and the University of Michigan.

DNA and RNA isolation and genotyping

Genomic DNA and RNA were prepared from liver tissue samples using either nuclei lysis buffer or Trizol respectively (Life Technologies, Carlsbad, CA) as previously described [15]. Briefly, following tissue lysis genomic DNA samples were phenol chloroform purified, ethanol washed, and resuspended in TE. RNA isolation used SpinSmart Columns (Denville, Metuchen, NJ) with DNaseI treatment to prevent contamination of genomic DNA in cDNA synthesis. cDNA was prepared using RNA and Reverse Transcriptase SSIII (Invitrogen, San Francisco, CA, USA), with controls lacking reverse transcriptase or template to test for residual gDNA contamination. SNPs in *CES1* were genotyped in liver samples using a primer extension assay (SNaPshot, Life Technologies), GC-clamp real time-PCR assay or fluorescently labeled PCR-restriction fragment length polymorphism (RFLP) analysis as previously described [21, 22]. PCR conditions and primers for all PCR assays (including genotyping) are provided in Supplemental Digital Content, Table 1.

Analysis of the CES1 Promoter Region

Initially Sanger sequencing was used to screen the immediate 5′UTR, exon 1 and intron 1 of CES1 to confirm the *CES1VAR* genetic sequence. Following association of *CES1VAR* with changes in mRNA and protein expression, the *CES1* promoter and upstream region (4.387

kb) were sequenced to scan for regulatory variants. This region included the annotated 5′UTR of *CES1,* through intron 1. PCR products were separated on agarose gel and the products mixed and fragmented on a Covaris S220 (Covaris, Woburn, MA). NEB Next Fast DNA Library Prep Set for Ion Torrent (New England Biosciences, Ipswich, MA) was used to generate barcoded libraries that were run on an Ion Torrent PGM (Life Technologies). Sequencing results were analyzed in CLC Bio's Genomics Workbench (CLC Bio, Katrinebjerg, Denmark), and SNP calling required, at minimum, 20 reads at a given base with 20% reads per allele.

The length of CES1 5′UTR (s) was measured using 5′RACE product synthesized using the First Choice RLM-RACE RNA Ligase Mediated RACE kit (Life Technologies) from a *CES1/CES1* individual. *CES1* 5′UTR length was visualized on an ABI3730 using product amplified with a *CES1* specific fluorescent tagged primer (Table 1, Supplemental Digital Content). Length was determined using the internal size standard Gene Scan 500 ROX (Life Technologies) on the ABI3730.

Analysis of *CES1* mRNA folding structure

Structure (2D) and free energy was assessed using mfold [23], with sequence based on the intermediate length 5′ UTR (derived from 5'RACE analysis), CES1 exon, and the first 18 bases of pGL3-Basic downstream of the insertion site. Folding structures and free energy were recorded and compared between the various constructs by One-way ANOVA (GraphPad Prism). To best visualize the overall difference in the predicted structures, three dimensional models were generated using RNAFold and RNAComposer (Figure 3) [24].

Luciferase reporter gene assays

Luciferase assay vectors were prepared using pGL3-Basic (Promega, Madison WI) and a *CES1* promoter region-5′UTR-exon 1 region fragment (1327 bases, Figure 4A). The Infusion Cloning System (Clontech, Mountain View, CA) was used to insert the fragment at the Hind III multiple cloning site of pGL3-Basic. Completed reactions were transformed into XL-2 (Agilent, Santa Clara, CA) cells and plated on LB agar with carbenicillin. Individual clones were screened and plasmids isolated by the Zyppy Mini Prep kit (Zymo Research, Irvine, CA). Plasmids were screened for inserts by Hind III digestion and sequenced for confirmation of the insert present in frame with the luciferase gene. Positive clones with haplotypes of interest were re-subcloned into DH5α cells to avoid colonized effects. Three positive subclones per haplotype were then combined and cultured, and plasmid DNA prepared with a Qiagen MidiPrep Kit (Qiagen, Valencia, CA). In total four constructs were generated, including, *CES1, CES1VAR, CES1SVAR* and *CES1* with *CES1* promoter SNP rs3815583. Luciferase assays were performed in HepG2 cells with high endogenous expression of CES1 (www.biogps.org) [25]. Prior to transfection, cells were plated in 12-well plates and grown to ~60% confluence. Transfection plasmid mixes of CES1-pGL3B and pRL (Promega), a Renilla fluorescence vector (to serve as control to normalize transfection efficiency), were prepared in a ratio of 1 μg to 0.2 μg in OptiMem Media (Life Technologies). Plasmid mix was added to an equivalent OptiMem mix of transfection reagent FuGene HD (Promega) at 1μg DNA/3 μl reagent. Six hours post transfection, media containing antibiotics were added to transfected wells (Penicillin/

Streptomycin, Life Technologies). Following incubation for 24 hours, luciferase activity was measured by Dual-Glo Luciferase Assay System (Promega) on a Packard Fusion plate reader (PerkinElmer Life and Analytical Sciences, Shelton, CT). Luciferase activity was determined as a ratio of luciferase fluorescence over renilla fluorescence. Luciferase was measured in three experiments, each with three replicates and two measurements per replicate.

Determination of CES1 mRNA by quantitative real-time PCR

CES1 mRNA expression was measured using quantitative real time PCR (qRT-PCR) on a 7500 Fast Real Time PCR System (Applied Biosystems, Carlsbad, CA, USA). Amplification of *CES1* mRNA captured *CES1, CES1VAR, CES1SVAR, CES1P1* and *CES1P1VAR* (the latter two accounting for ~2% of total CES1 mRNA). Relative quantitation used *GAPDH* as a housekeeping gene where expression is representative of overall RNA quality in human liver and has low interindividual variability [26].

Measurements of gDNA and mRNA allelic ratios by primer extension (SNaPshot)

The SNaPshot protocol has been previously described [15, 22]. In short, PCR amplification surrounding a marker SNP in an exonic region, from a heterozygous sample, is visualized using an extension primer and fluorescent ddNTPs. Allelic mRNA ratios are determined on an ABI3730 sequencer (Life Technologies). The average of the allelic ratios of gDNA is used to normalize allelic mRNA ratios in each tissue. The primary marker SNP for *CES1* was rs12149370, derived from the translocation affected portion of *CES1VAR* and *CES1SVAR*. Because of *CES1's* high identity with the *CES1P1/CES1P1VAR* gene locus, the genomic DNA control PCR product was selected to be specific to the *CES1* locus regardless of translocation, providing accurate allelic gDNA ratios. Allelic mRNA ratios were measured at rs12149370 using PCR conditions detecting both CES1 (including CES1VAR) and CES1P1/CES1P1VAR, the latter expressed at low levels in the liver. qRT-PCR indicated that *CES1P1/CES1P1VAR* transcripts account for only ~2% of all CES1 mRNA, which was considered a negligible contribution to both allelic mRNA ratios and total mRNA. SNaPshot measurements were completed in duplicate for gDNA in all 28 livers heterozygous for rs12149370 (variant included in *CES1VAR* and *CES1SVAR*). Allelic mRNA ratios were measured in triplicate for 26 of 28 samples, while L36 and L120 were measured in duplicate. Primers for this assay are provided in Supplemental Digital Content, Table 1. Significant allelic mRNA ratios deviating from unity were considered greater than 1.2-fold and was termed allelic expression imbalance (AEI), an indicator of *cis*-acting regulatory variation.

Measurement of CES1 protein quantity and metabolic activity

Absolute CES1 protein expressions were determined in 102 individual human liver S9 fraction (HLS9) samples using a novel LC-MS/MS-based targeted absolute quantitative proteomics method (TAQSI) recently established in our laboratory [27]. Briefly, an aliquot of 20 μg protein of HLS9 samples was mixed with the internal standard SILAC HepG2 cell S9 fractions (40 μg protein). After reduction and alkylation, proteins were digested with trypsin (Worthington Biochemical Corporation, Freehold, NJ) at an enzyme/protein ratio of 1:500 in an incubation shaker at 200 rpm at 37°C for 16 h. Digested peptides were then

extracted (Oasis HLB columns, Waters, Milford, MA) and vacuum dried (Speed Vac SPD1010, Thermo Scientific, Hudson, NH). Finally, samples were reconstituted in 50 μl 50% acetonitrile and subjected to LC-MS/MS analysis. Standard calibration curves established from unlabeled CES1 protein calibrators (R&D system, Minneapolis, MN) and the HepG2 SILAC internal standards were employed to quantify absolute CES1 protein expressions in individual human livers. The CES1 calibrators ranged from 0.59 to 11.8 pmol, which the covered normal range of CES1 expressions in 20 μg proteins of HLS9 preparations. Three quality control (QC) samples (1.18, 2.36 and 4.73 pmol recombinant CES1) were utilized to evaluate the accuracy and precision of the assay.

To determine CES1 catalytic activity in the 102 individual human livers, the CES1 substrates clopidogrel (300 μM), enalapril (500 μM), or methylphenidate (50 μM) were incubated with individual HLS9 samples at 37°C for 20, 10, and 60 minutes, respectively. The protein concentrations of HLS9 for the three substrates were 1.5, 0.5, 0.5 mg/ml, respectively. Preliminary experiments were performed to ensure the formation of hydrolytic metabolites was linear with CES1 concentrations in the HLS9 samples. Following incubation, the concentrations of the metabolites clopidogrel carboxylate, enalaprilat, and ritalinic acid were measured using the LC-MS/MS assays previously established in our laboratory [10, 28, 29]. All incubations were carried out in triplicate.

Data Analysis

Association of *CES1VAR* with gene expression and protein quantity was tested in SPSS (IBM, Armonk, NY). Stepwise linear regression including covariates, gender, race, rs3815583 carrier status, rs2244613 carrier status, and *CES1VAR* carrier status were included in these analyses. T-tests, correlation statistics for RNA-Protein relationship and one-way ANOVA for luciferase activity were performed in GraphPad Prism. P-values < 0.05 were considered significant for all associations. Allele frequencies for *CES1VAR* and other variants of interest were assessed using Helix Tree SNP and Variation Suite (Golden Helix, Bozeman, MT, USA) (Table 2B).

Results

An alternative 5′ variant of CES1 and 5′UTR Usage

To understand the architecture of *CES1* and its relationship to *CES1P1*, Sanger sequencing of the *CES1* gene locus was performed. Previous studies have characterized the gene loci of *CES1* and *CES1P1* and identified *CES1VAR* (where a 5′ portion of *CES1P1* translocates to *CES1,* MAF = 0.17, Table 2B) [9]. Sequencing gDNA from *CES1/CES1, CES1/CES1VAR* and *CES1VAR/CES1VAR* homozygotes defined the major regions affected by *CES1VAR*. The *CES1VAR* 5′ boundary begins between 5′UTR SNP rs3815583 (outside of the region affected by translocation) and rs12149373 (first SNP of *CES1VAR*), and ends in intron 1. In total, this includes 11 variants in complete linkage disequilibrium (LD) attributable to *CES1VAR* in place of the original *CES1* sequence (Figure 2). When screening for variants using *CES1* specific primers, 3 samples were identified with 5 *CES1P1VAR* SNPs in the 5′UTR of *CES1*, while the exon 1 coding regions and intron 1 were unaltered. This shorter

translocation variant is referred to here as *CES1SVAR*, representing a rare 5′ sequence variation seen at the *CES1* gene locus (MAF < 0.01 Table 2B).

To select the most abundant 5′ mRNA isoforms for allelic mRNA analysis, *CES1* 5′UTR length was analyzed by 5′ RACE. In a *CES1* homozygote, fluorescently labeled CES1-RACE PCR products indicate that CES1 mRNA in human liver is largely ~ 50 bases shorter than the mRNA annotated by RefSeq and the UCSC Genome Browser (Figure 1, Supplemental Digital Content), in relative concordance with previously reported transcription start sites [17]. qRT-PCR confirms this medium-length 5′UTR length mRNA as the primary species, while the annotated long 5′UTR accounts for ~2% of total CES1 expression (including *CES1, CES1, CES1P1* and *CES1P1VAR data not shown)*. A shorter mRNA isoform beginning at the exon 1 coding region was also detectable with the RACE analysis, but only in small quantities (Figure 1, Supplemental Digital Content). Based on these observations, the medium 5′UTR is an abundant mRNA isoform beginning just downstream from *CES1VAR* and *CES1SVAR* associated sequence. Notably, this medium 5′UTR causes the *CES1VAR* variant rs12149373 to exist in the immediate promoter region, potentially affecting gene expression.

RNA folding in the 5′ region of *CES1*

Considering the variants associated with *CES1, CES1VAR* and *CES1SVAR,* and the abundance of the medium–length 5′UTR, representative 5′UTR/exon1 sequences were analyzed for differences in RNA folding using mfold. Sequence variation can be viewed in Figure 2, starting at the beginning of the medium 5′UTR, through exon 1. The CES1 sequence showed 2 folding patterns, with an average final ΔG = −27.74. The CES1VAR sequence yielded 4 folding patterns with an average ΔG = −40.58. The CES1SVAR sequence yielded 3 structures with an average ΔG = −40.68. When compared to CES1, the CES1VAR and CES1SVAR sequences had significantly lower free energy ($P < 0.001$, One-way ANOVA). Though there is some overlap in the predicted structures, the difference in ΔG indicates that CES1VAR and CES1SVAR are more stable in their predicted structures when compared to CES1. RNAFold outputs confirmed mfold free energy predictions. The RNAFold 2-D structure was visualized in 3-D using RNA Composer, with drastically different 3-D structures between the CES1, CES1VAR and CES1SVAR sequences (Figure 3). These indicate overall structural differences that may affect both transcription and translation.

Luciferase reporter assays of the *CES1* 5′ region in HepG2

To determine the function of the various *CES1* 5′ variant alleles, luciferase assays were performed in HepG2 cells. Constructs contained complete 5′UTR and exon 1 regions corresponding to *CES1, CES1VAR, CES1SVAR* and *CES1*+ rs3815583 (Figure 4A and Table 2, Supplemental Digital Content), with an additional 1172 bases of promoter region. *CES1VAR* showed a significant 35% decrease in luciferase activity when compared to *CES1* ($P < 0.01$, at 24 hrs) (Figure 4B). In contrast, *CES1SVAR* showed an average 36% increase in luciferase activity when compared to *CES1* ($P < 0.01$), though there was considerable variation in activity between replicates. *CES1SVAR* was not further studied because of its low MAF. Presence of rs3815583 (a SNP specific to *CES1*) showed no significant difference

in luciferase activity. The difference between *CES1* and *CES1VAR* luciferase activity was confirmed by measuring activity after 48 hours. Here, *CES1VAR* luciferase activity is significantly decreased by 22% (P = 0.012, Figure 2, Supplemental Digital Content).

CES1VAR associates with decreased total mRNA expression

Previous studies have shown that *CES1* gene expression is highly variable, while the contributions of CES1 variants were unknown. Total CES1 mRNA was measured in 60 human liver samples (*CES1* = 29, *CES1/CES1VAR* = 24, *CES1VAR/CES1VAR* =7). Analysis by stepwise linear regression including gender, race, and the previously reported rs3815583 and rs224816, indicated that *CES1VAR* carrier status was the only significant contributor to gene expression (p = 0.003; *CES1SVAR* was too infrequent and not included). Even with substantial variability in mRNA levels, mean total CES1 expression was 2.6-fold lower in *CES1VAR* carriers (heterozygotes and homozygotes) than in non-carriers (p =0.015, two-tailed t-test, WT; n = 29; VAR; n = 31) (Figure 5).

CES1VAR is associated with allelic expression imbalance (AEI)

Allelic mRNA ratios were measured to determine whether CES1 is under the influence of *cis*-acting regulatory polymorphism(s). Because of the minor contribution of *CES1P1/CES1P1VAR* to total gene expression, allelic mRNA ratios were measured using common CES1 primers, amplifying almost exclusively CES1 mRNA. Allelic mRNA ratios were screened in 28 human livers heterozygous for the 5′ UTR rs12149370 located in the abundant medium-length mRNA transcript of both *CES1VAR* (minor allele frequency = 0.17) and *CES1SVAR* (MAF < 0.01). Different PCR primers were used for amplifying gDNA (to distinguish between *CES1* and *CES1P1* gene loci) and mRNA. To account for variability of allelic gDNA and mRNA ratios, and differences in amplification efficiency between reactions, the threshold for observing AEI was set at >1.20-fold. Of the 28 livers, 18 display AEI ratios >1.2-fold (1.23 to 1.4-fold, Figure 6), indicating that the *CES1VAR* allele was expressed at a lower level compared to the *CES1* allele, and that *cis*-acting regulatory polymorphisms influence CES1 expression. *CES1SVAR* carriers, L50, L51, and LL10 show some degree of AEI (2 of 3 >1.2 fold), with similar interindividual variability observed in *CES1VAR* carriers. In 10 livers heterozygous for either translocation, the allelic mRNA ratios did not reach significance or were close to unity. This result could have been caused by variability in the allelic ratio analysis (4 samples close to the cutoff), variability in the magnitude of the allele effect caused by trans-acting factors, presence of additional regulatory variants compensating for the *CES1VAR* effect, or an alternative truly functional regulatory variant in partial LD with *CES1VAR* –though no variants in high LD and with the requisite frequency are present in the 1000 genomes database. In some cases, allelic mRNA ratios greater than 2-fold were observed and may be the result of an uncommon variant (representative sample L18, Fig. 6).

Search for additional variants in the *CES1* promoter region

To test whether AEI was associated with a *cis*-acting functional polymorphism other than *CES1VAR* or *CES1SVAR*, 4.387 kb of the promoter region was sequenced in 10 human livers (5 AEI positive and 3 AEI negative *CES1/CES1VAR* individuals, and 1 AEI positive and 1 AEI negative *CES1/CES1SVAR* individuals). No single SNP was significantly

associated with the allelic mRNA ratios. Promoter SNP rs38135583 was present in all AEI-negative samples (n = 4), and present in only 1 of 6 AEI-positive samples. This SNP is located in the long low-abundance 5′UTR isoform of *CES1* and is not in LD with *CES1VAR* though it has been associated with appetite reduction during methylphenidate therapy [12]. Analysis in additional liver cohorts indicate rs38135583 was not associated with AEI status (n=28) or mRNA expression (n=60).

CES1VAR association with CES1 protein expression

The association of *CES1VAR* with decreased CES1 mRNA expression prompted us to determine allelic effects at the protein level. The novel targeted proteomics assay TAQSI was found to be highly accurate and precise for absolute quantification of CES1 expression in liver. The inter- and intra-day precision measured as relative standard deviations in the three quality controls were equal or less than 9.0% and 5.5%, respectively. The inter- and intra-day accuracy were within the ranges of 96.1%–102.2% and 94.6%–106.1%, respectively. The distributions of *CES1* genotypes among those samples are: *CES1SVAR* heterozygote: 1; *CES1VAR* homozygotes: 6; *CES1VAR* heterozygotes: 32, *CES1VAR* non-carriers: 63. The CES1 protein expression levels in *CES1VAR* carriers did not significantly differ from those of non-carriers (161 ± 63 vs 180 ± 76 pmole/mg protein, p = 0.24) (Figure 7A). In addition, we measured CES1 mRNA expression in 50 livers randomly selected from the whole set of 102 samples, and found that CES1 protein expressions did not significantly correlate to CES1 mRNA levels (Figure 7B, p = 0.67, r = 0.06, n = 50).

CES1VAR association with metabolic activity

The effect of *CES1VAR* on CES1 enzymatic activities was further interrogated using a total of 102 HLS9 samples. The hydrolyzing velocity of the CES1 substrates, enalapril, clopidogrel and methylphenidate, was measured and tested for association with *CES1VAR* carrier status by linear regression including gender, race, as well as the previously reported variants, including rs3815583 and rs2244613. No predictor was significantly associated with enzyme activity. Hydrolysis of enalapril in *CES1VAR* carriers was not significantly decreased when compared to the non-carriers (p = 0.93, two-tailed t-test, *WT*: 53 ± 38 (pmole/min/mg protein, n = 63) vs. *VAR*: 54 ± 37 (pmole/min/mg protein, n = 38)) (Figure 8A). Similar results were obtained when comparing CES1 activities between the *CES1VAR* carriers and non-carriers on the metabolisms of clopidogrel (p = 0.82, two-tailed t-test, *WT*: 130 ± 58 (pmole/min/mg protein, n =63) *vs. VAR*: 137 ± 64 (pmole/min/mg protein, n = 38)) (Figure 8B) and methylphenidate (p = 0.95, two-tailed t-test, *WT*: 205 ± 98 (pmole/hr/mg protein, n =63) *vs. VAR*: 208 ± 95 (pmole/hr/mg protein, n = 38)) (Figure 8C). Notably, G143E, a well defined loss of function allele was significantly associated with CES1 activity on hydrolysis for all three CES1 substrates (*data not pictured)*.

Discussion

This study confirms previous reports of the *CES1P1-CES1* translocation generating the *CES1VAR* allele (17% MAF) with 11 SNPs in the 5′UTR, exon1 and intron 1 derived from the *CES1P1* sequence. In addition we identified a shorter translocation involving only the 5′UTR, termed *CES1SVAR*, with MAF < 0.01 (Figure 2). As *CES1P1* is poorly expressed

and accounts for only ~2% of the CES1 mRNA in the liver, we tested whether the translocation affects expression of CES1 mRNA. Though CES1 mRNA was variable in human liver, results of this study indicate that *CES1VAR* decreases mRNA expression in human liver to a moderate extent which is supported by results from a luciferase reporter gene assay. Similarly, *CES1SVAR* displayed only a modest effect in a luciferase reporter gene assay, but was not further studied because of its low frequency. However, the studies of CES1 protein expression and metabolic activity on hydrolyzing enalapril, clopidogrel, and methylphenidate in human livers failed to reveal significant differences between the *CES1VAR* genotypes. Overall, our results suggest a modest reduction in CES1 mRNA expression caused by the translocation in the *CES1VAR* allele. Such small effects could be buffered by large interindividual variability in CES1 expression and activity, as well as possible post-transcriptional regulatory processes. As a result, any effect of *CES1VAR* on CES1 protein expression and activity might not be readily detected. The results are consistent with a recent study by Wang and colleagues, suggesting that *CES1VAR* does not significantly affect hepatic CES1 protein expression and enzymatic activity on the activation of ACE inhibitor prodrugs [28].

Reduced expression due to *CES1VAR* is supported by total and allelic mRNA expression data, and luciferase reporter gene analysis. Our allelic mRNA expression results are consistent with this finding but leave open the possibility of additional regulatory variants. A previous study had suggested that less common exon 1 *CES1* regulatory variants exist in addition to *CES1VAR* [18]. Although we identified *CES1SVAR* by sequencing, allelic mRNA ratios failed to reveal any new regulatory variants in the screened 5′ region of *CES1*. Our study also provides clarification on *CES1* 5′UTR length. Measurement of the average 5′UTR length by 5′RACE indicates that the most frequent species is shorter than the RefSeq annotation. This medium-length 5′UTR is in concordance with transcription start sites reported previously for *CES1* [17].

Luciferase assays support 5′UTR variation as a functional variant, indicating that *CES1VAR* decreases protein expression one-third in HepG2 cells, while *CES1SVAR* affected luciferase activity with a one-third increase. Construct *CES1* + rs3815583 (5′ UTR SNP) includes a variant only present in the long 5′ UTR transcript. While this SNP has been associated with methylphenidate side effects the construct showed no effect on luciferase activity.

The sequences of *CES1VAR* and *CES1SVAR* introduce multiple factors that may affect expression. First, the medium 5′UTR begins just downstream from the first *CES1VAR* and *CES1SVAR* associated variant, potentially affecting transcription. In the exonic coding region, *CES1VAR* contains multiple non-synonymous coding variants and includes several intron 1 SNPs, all with unknown effects. Second, in silico analysis of *CES1VAR* and *CES1SVAR* reveals changes in computed folding structure of the mRNA, with *CES1* displaying higher free energy when compared to both variant alleles. Here, significant decreases in free energy may cause the variant allele's mRNA to be less receptive to translation. These factors are all potential mechanisms by which *CES1VAR* and *CES1SVAR* could impact transcription, mRNA processing, and translation at the *CES1* gene locus – some effects potentially cancelling each other out. Interestingly, the poorly expressed *CES1P1* specific sequence which creates *CES1VAR* might also be expected to drastically

decrease *CES1VAR* expression in a similar manner. However, we see only a small *CES1VAR* specific effect at the genetic level, with no effect at the protein level. Despite the possible regulatory elements contained in the translocated region, this indicates that more pervasive regulatory elements are present outside the translocation.

Here we have shown that the previously described *CES1VAR* decreases CES1 mRNA expression in the human liver by ~30%, possibly reflecting complex regulation of the 5′ region. Multiple UTR lengths, alternative sequences, and potential changes in mRNA folding suggest this region is driving variation in CES1 expression. However, the associations of *CES1VAR* genotypes with CES1 protein expression and activity in human livers were insignificant in the present study; any small effect could have been confounded by the marked interindividual variability, while we also cannot exclude possible involvement of post-translation regulation in CES1 protein expression. Future studies should assess the effect of *CES1* regulatory variants on the metabolism of CES1 substrate drugs in clinical settings to evaluate the utility of those variants as clinical biomarkers to improve therapeutic outcomes of many drugs metabolized by CES1.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: This study was supported by a grant from the NIH General Medical Sciences U01 GM092655, and partially by a grant from the National Center for Research Resources (UL1RR025755). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Research Resources or the National Institutes of Health.

References

1. Lins R, Broekhuysen J, Necciari J, Deroubaix X. Pharmacokinetic profile of 14C-labeled clopidogrel. Semin Thromb Hemost. 1999; 25:29–33. [PubMed: 10440420]
2. Suzaki Y, Uemura N, Takada M, Ohyama T, Itohda A, Morimoto T, et al. The effect of carboxylesterase 1 (CES1) polymorphisms on the pharmacokinetics of oseltamivir in humans. Eur J Clin Pharmacol. 2013; 69:21–30. [PubMed: 22673926]
3. Shi D, Yang J, Yang D, LeCluyse EL, Black C, You L, et al. Anti-influenza prodrug oseltamivir is activated by carboxylesterase human carboxylesterase 1, and the activation is inhibited by antiplatelet agent clopidogrel. J Pharmacol Exp Ther. 2006; 319:1477–1484. [PubMed: 16966469]
4. Sun Z, Murry DJ, Sanghani SP, Davis WI, Kedishvili NY, Zou Q, et al. Methylphenidate is stereoselectively hydrolyzed by human carboxylesterase CES1A1. J Pharmacol Exp Ther. 2004; 310:469–476. [PubMed: 15082749]
5. Humerickhouse R, Lohrbach K, Li L, Bosron WF, Dolan ME. Characterization of CPT-11 hydrolysis by human liver carboxylesterase isoforms hCE-1 and hCE-2. Cancer Res. 2000; 60:1189–1192. [PubMed: 10728672]
6. Song JC, White CM. Clinical pharmacokinetics and selective pharmacodynamics of new angiotensin converting enzyme inhibitors: an update. Clin Pharmacokinet. 2002; 41:207–224. [PubMed: 11929321]
7. Thomsen R, Rasmussen HB, Linnet K. Consortium TI. In Vitro Drug Metabolism by Human Carboxylesterase 1: Focus on Angiotensin-Converting Enzyme Inhibitors. Drug Metab Dispos. 2014; 42:126–133. 2014. [PubMed: 24141856]

8. Geshi E, Kimura T, Yoshimura M, Suzuki H, Koba S, Sakai T, et al. A single nucleotide polymorphism in the carboxylesterase gene is associated with the responsiveness to imidapril medication and the promoter activity. Hypertens Res. 2005; 28:719–725. [PubMed: 16419644]

9. Fukami T, Nakajima M, Maruichi T, Takahashi S, Takamiya M, Aoki Y, et al. Structure and characterization of human carboxylesterase 1A1, 1A2, and 1A3 genes. Pharmacogenet Genomics. 2008 Oct.18:911–920. [PubMed: 18794728]

10. Zhu HJ, Patrick KS, Yuan HJ, Wang JS, Donovan JL, DeVane CL, et al. Two CES1 gene mutations lead to dysfunctional carboxylesterase 1 activity in man: clinical significance and molecular basis. Am J Hum Genet. 2008; 82:1241–1248. [PubMed: 18485328]

11. Lewis JP, Horenstein RB, Ryan K, O'Connell JR, Gibson Q, Mitchell BD, et al. The functional G143E variant of carboxylesterase 1 is associated with increased clopidogrel active metabolite levels and greater clopidogrel response. Pharmacogenet Genomics. 2013; 23:1–8. [PubMed: 23111421]

12. Bruxel EM, Salatino-Oliveira A, Genro JP, Zeni CP, Polanczyk GV, Chazan R, et al. Association of a carboxylesterase 1 polymorphism with appetite reduction in children and adolescents with attention-deficit/hyperactivity disorder treated with methylphenidate. Pharmacogenomics J. 2013; 13:476–480. [PubMed: 22688218]

13. Johnson KA, Barry E, Lambert D, Fitzgerald M, McNicholas F, Kirley A, et al. Methylphenidate side effect profile is influenced by genetic variation in the attention-deficit/hyperactivity disorder-associated CES1 gene. J Child Adolesc Psychopharmacol. 2013; 23:655–664. [PubMed: 24350812]

14. Pare G, Eriksson N, Lehr T, Connolly S, Eikelboom J, Ezekowitz MD, et al. Genetic determinants of dabigatran plasma levels and their relation to bleeding. Circulation. 2013; 127:1404–1412. [PubMed: 23467860]

15. Wang D, Guo Y, Wrighton SA, Cooke GE, Sadee W. Intronic polymorphism in CYP3A4 affects hepatic expression and response to statin drugs. Pharmacogenomics J. 2011; 11:274–286. [PubMed: 20386561]

16. Wang D, Poi MJ, Sun X, Gaedigk A, Leeder JS, Sadee W. Common CYP2D6 polymorphisms affecting alternative splicing and transcription: long-range haplotypes with two regulatory variants modulate CYP2D6 activity. Hum Mol Genet. 2014; 23:268–278. [PubMed: 23985325]

17. Hosokawa M, Furihata T, Yaginuma Y, Yamamoto N, Watanabe N, Tsukada E, et al. Structural organization and characterization of the regulatory element of the human carboxylesterase (CES1A1 and CES1A2) genes. Drug Metab Pharmacokinet. 2008; 23:73–84. [PubMed: 18305377]

18. Tanimoto K, Kaneyasu M, Shimokuni T, Hiyama K, Nishiyama M. Human carboxylesterase 1A2 expressed from carboxylesterase 1A1 and 1A2 genes is a potent predictor of CPT-11 cytotoxicity in vitro. Pharmacogenet Genomics. 2007; 17:1–10. [PubMed: 17264798]

19. Sai K, Saito Y, Tatewaki N, Hosokawa M, Kaniwa N, Nishimaki-Mogami T, et al. Association of carboxylesterase 1A genotypes with irinotecan pharmacokinetics in Japanese cancer patients. Br J Clin Pharmacol. 2010; 70:222–233. [PubMed: 20653675]

20. Yamada S, Richardson K, Tang M, Halaschek-Wiener J, Cook VJ, Fitzgerald JM, et al. Genetic variation in carboxylesterase genes and susceptibility to isoniazid-induced hepatotoxicity. Pharmacogenomics J. 2010; 10:524–536. [PubMed: 20195289]

21. Wang D, Johnson AD, Papp AC, Kroetz DL, Sadee W. Multidrug resistance polypeptide 1 (MDR1, ABCB1) variant 3435C>T affects mRNA stability. Pharmacogenet Genomics. 2005; 15:693–704. [PubMed: 16141795]

22. Pinsonneault J, Nielsen CU, Sadee W. Genetic variants of the human H+/dipeptide transporter PEPT2: analysis of haplotype functions. J Pharmacol Exp Ther. 2004; 311:1088–1096. [PubMed: 15282265]

23. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003; 31:3406–3415. [PubMed: 12824337]

24. Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, et al. Automated 3D structure composition for large RNAs. Nucleic Acids Res. 2012; 40:e112. [PubMed: 22539264]

25. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. 2009; 10:R130. [PubMed: 19919682]

26. Barber RD, Harmer DW, Coleman RA, Clark BJ. GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. Physiol Genomics. 2005; 21:389–395. [PubMed: 15769908]

27. Wang X, Liang Y, Liu L, Shi J, Zhu HJ. Targeted absolute quantitative proteomics with SILAC internal standards and unlabeled full-length protein calibrators (TAQSI). Rapid Commun Mass Spectrom. 2016 In Press.

28. Wang X, Wang G, Shi J, Aa J, Comas R, Liang Y, et al. CES1 genetic variation affects the activation of angiotensin-converting enzyme inhibitors. Pharmacogenomics J. 2015

29. Zhu HJ, Wang X, Gawronski BE, Brinda BJ, Angiolillo DJ, Markowitz JS. Carboxylesterase 1 as a determinant of clopidogrel metabolism and activation. J Pharmacol Exp Ther. 2013; 344:665–672. [PubMed: 23275066]

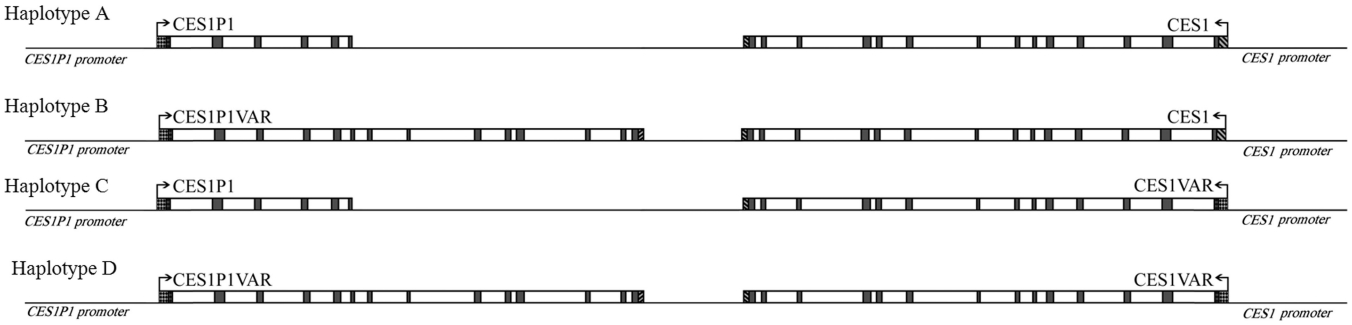

Figure 1. *CES1* gene loci on chromosome 16

CES1, CES1P1 and associated structural variants exist in a tail to tail configuration in the displayed haplotypes. *CES1P1* can exist as either the pseudogene or the full length functional *CES1P1VAR. CES1P1* sequence in the 5′UTR, exon 1 and intron 1 is translocated as a group of 11 highly linked SNPs to the *CES1* gene locus creating the *CES1VAR* isoform observed in Haplotypes C and D. Striped or diamond patterns indicate regions containing variation specific to either *CES1* or *CES1P1* respectively, and grey boxes indicate coding exons. Fine detail on sequence differences between isoforms is shown in Figure 2 and Table 2.

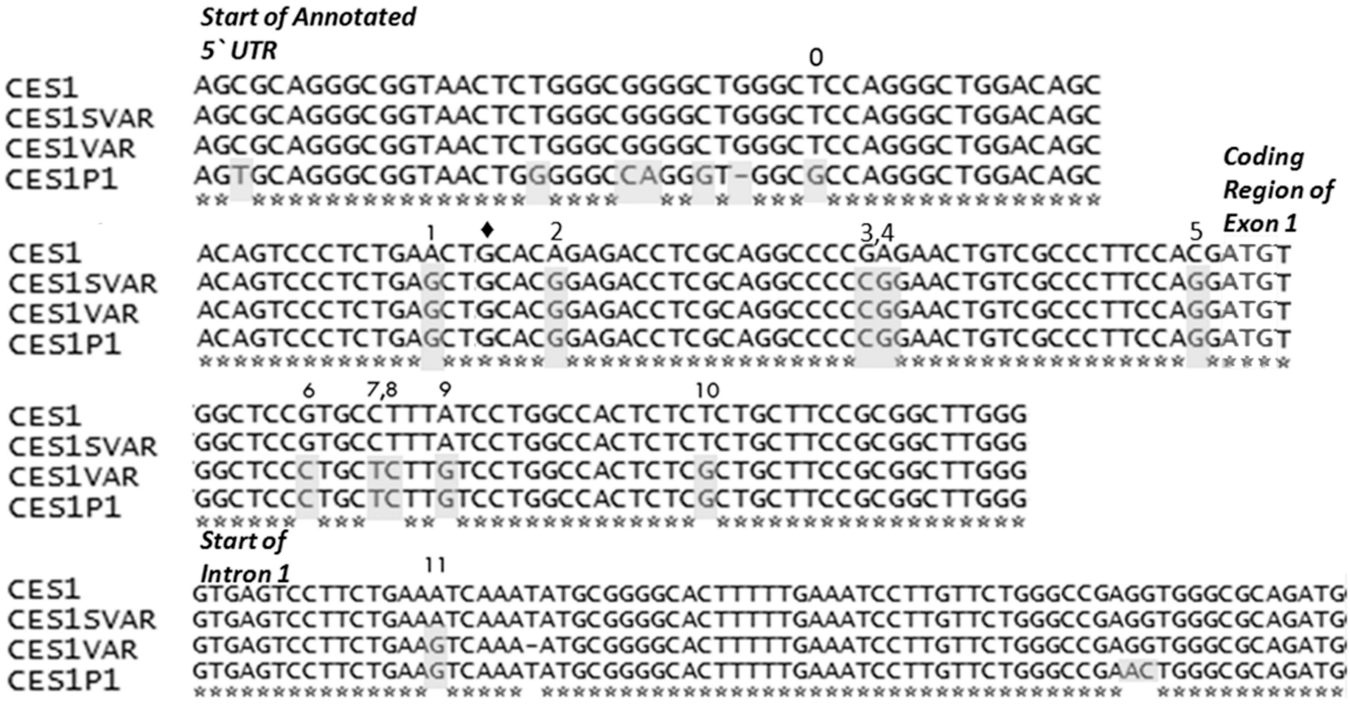

Figure 2. Fine detail of CES1 isoforms, *CES1, CES1VAR, CES1SVAR* and *CES1P1/CES1P1VAR.*

Polymorphic bases are highlighted in grey while those included in the translocation are numbered. *CES1VAR* includes variants 1–11, while *CES1SVAR* includes only the 5′ UTR variants 1–5. Not included in the translocation(s) SNP "0" is not in strong linkage with SNPs 1–11, and thusly can occur in *CES1/CES1* individuals. Sequence begins at the start of the annotated 5′ UTR, while the start of the medium length 5′ UTR is indicated by the ♦. Notably, SNP "1" is just upstream from the start of the medium length 5′ UTR. The start of the exon 1 coding region is indicated in text as are the end of exon 1 and start of intron 1. SNP numbers correspond to "rs" numbers in Table 2A.

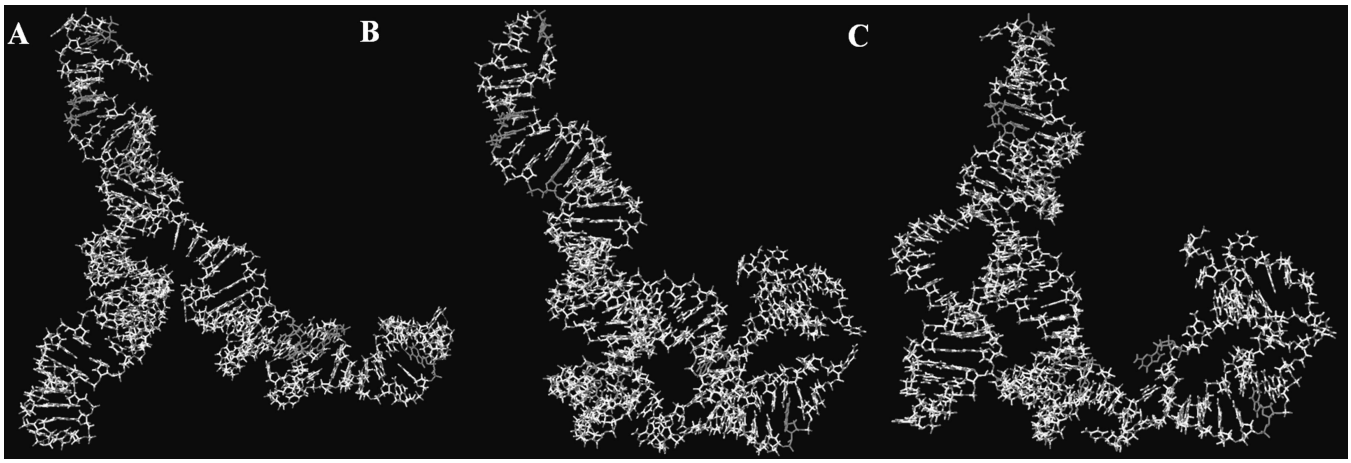

Figure 3. 3D Structures of *CES1* 5′ UTR isoforms

Three dimensional structures are based on the medium *CES1* 5′UTR and a portion of the following pGL3-Basic vector used in the luciferase assay. Structures for *CES1* (A), *CES1VAR* (B) and *CES1SVAR* (C) include dark-grey dNTPs which represent positions where variation occurs between isoforms. In all cases *CES1* exonic region differences are located in the upper helix, while 5′ region variation is in the lower region.

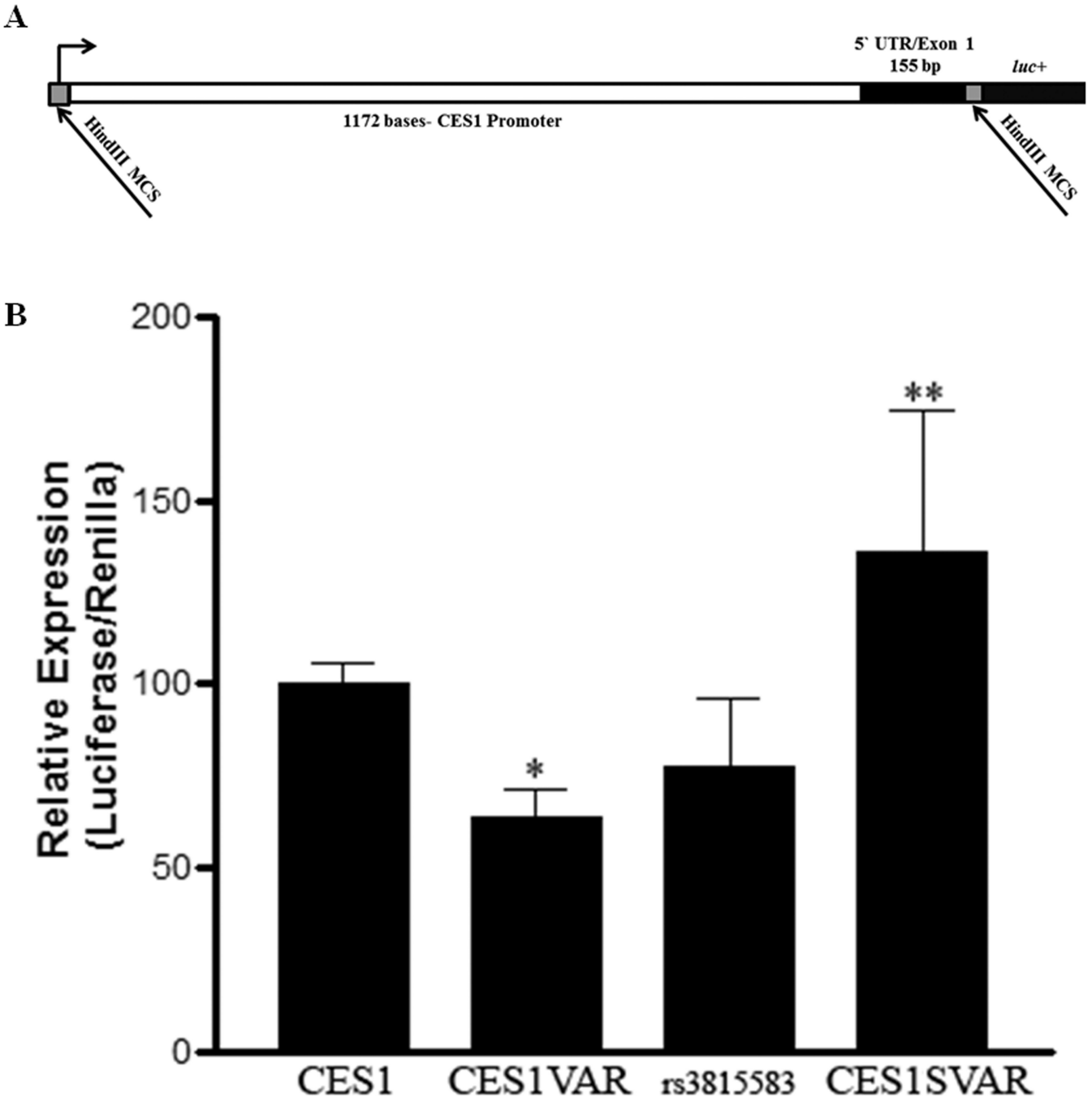

Figure 4. Luciferase activity of *CES1* constructs in HepG2 cells

(A) Construct schematic, specific construct differences can be viewed in Supplemental Digital Content, Table 2.

(B) *CES1VAR* luciferase activity was significantly lower than *CES1* (35%, P <0.01). *CES1SVAR* was significantly higher than CES1WT activity (36%, P <0.01). Data are presented as the average of 3 transfection experiments. Significant differences when compared to WT are indicated by * (*CES1VAR)* and ** (*CES1SVAR).*

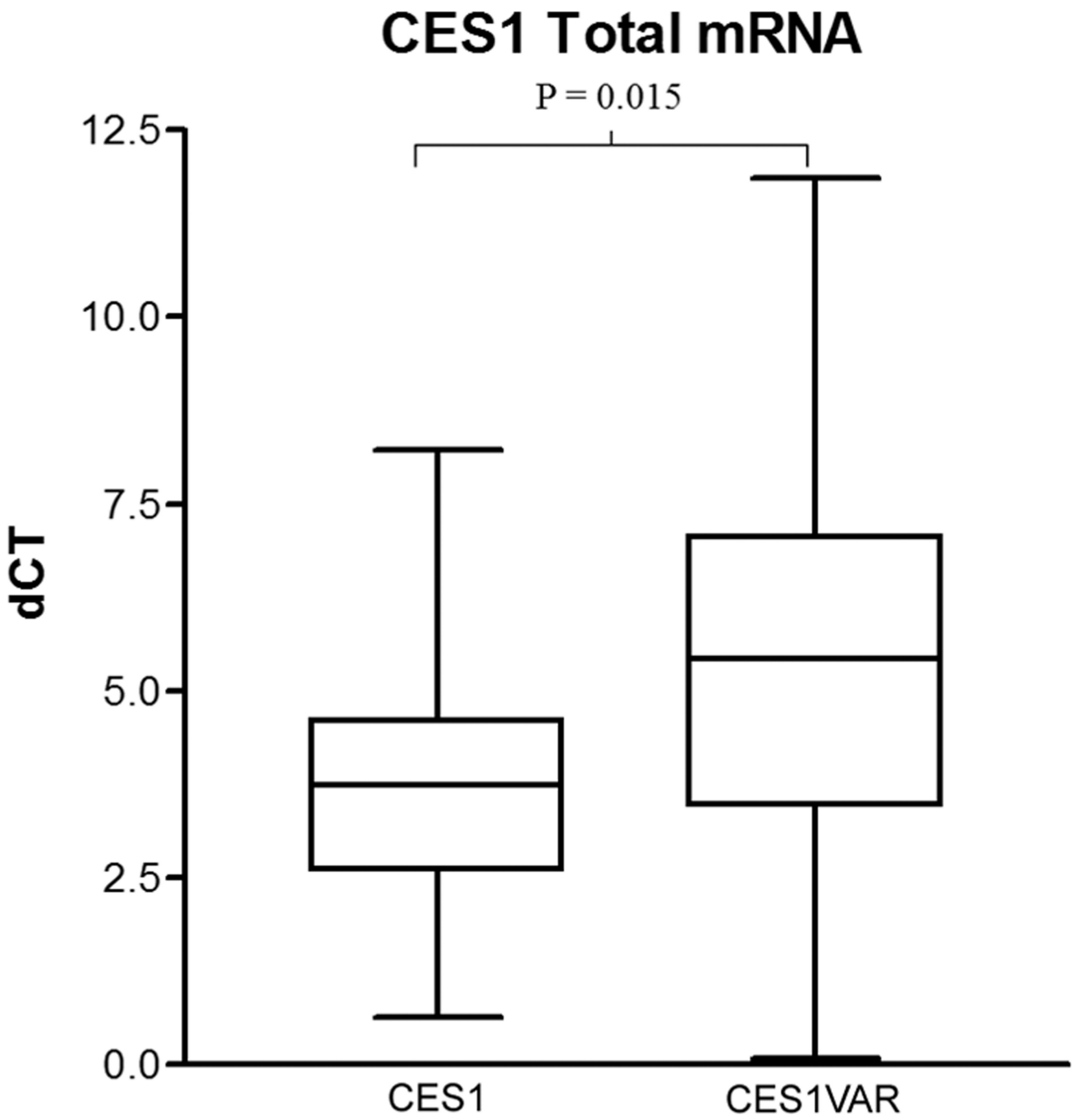

Figure 5. *CES1VAR* associates with decreased CES1 gene expression

CES1VAR carrier status was the only significant predictor of gene expression when analyzed with gender, race, rs2244613 and rs3815583 as covariates (P = 0.003). *CES1VAR* carrier status was associated with decreased mRNA expression (sum of CES1 and CES1VAR mRNA, WT, n = 29; VAR, n = 31). mRNA levels are 2.6-fold (mean, p = 0.015) lower in carriers. Data are presented as a box plot and whisker plot where the box shows median with the 25th and 75th percentiles while minimum and maximum values are shown by whiskers.

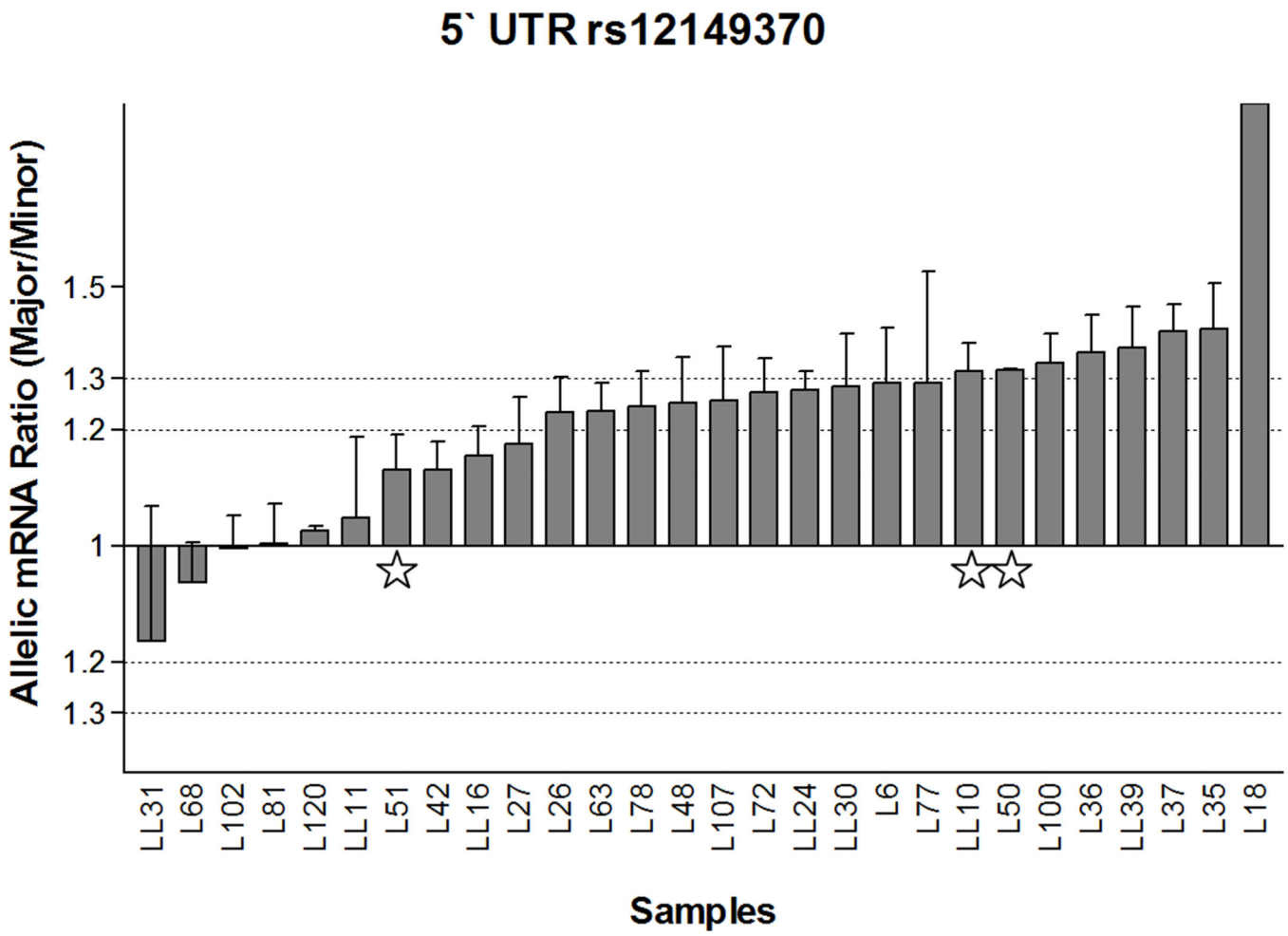

Figure 6. Allelic CES1 mRNA expression at rs12149370 in livers

The allelic RNA ratio is expressed as a ratio of major/minor alleles, corrected for gDNA allelic ratios. The marker rs12149370 is part of both *CES1VAR* and *CES1SVAR* alleles. *CES1SVAR* individuals are marked with a star, while all other samples are *CES1VAR* heterozygotes. Allelic expression is considered significant when >1.2-fold. AEI-positive samples have an average of ~1.35-fold allelic mRNA ratio (18 of 28 samples). Data are mean±SD.

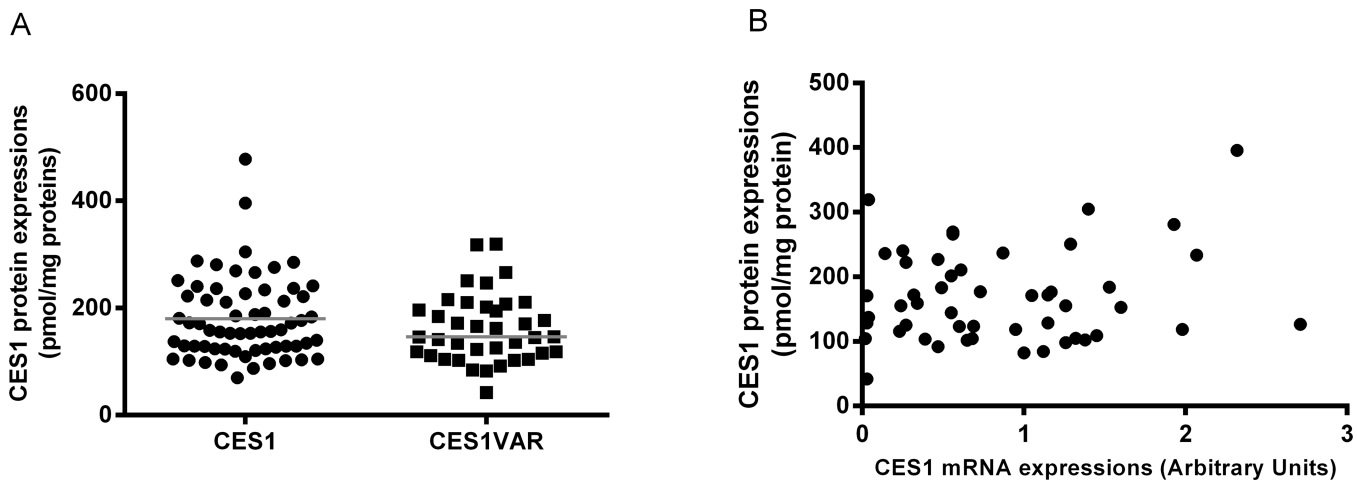

Figure 7. CES1 protein quantity is not associated with *CES1VAR* genotypes, No correlation between CES1 protein and RNA

(A) Absolute CES1 protein levels were not associated with *CES1VAR* genotypes in human livers. Horizontal lines indicate sample mean.

(B) CES1 mRNA and protein level do not significantly correlate with one another (p=0.67, r = 0.06, n = 50).

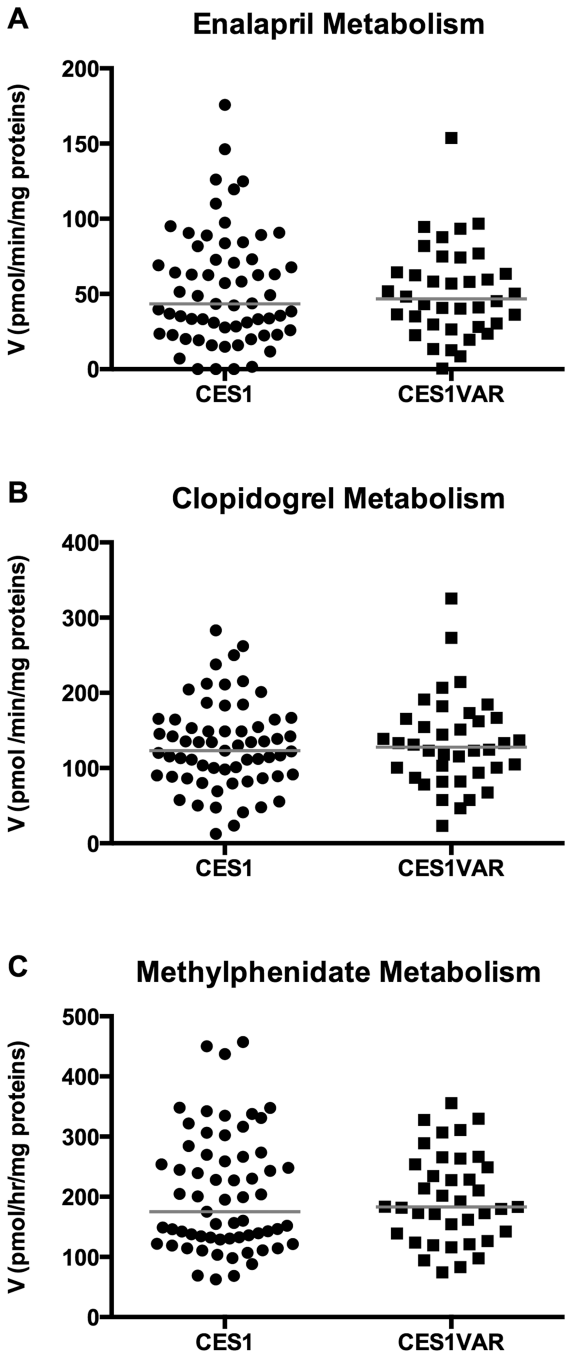

Figure 8. CES1VAR does not associate with metabolic activity

CES1VAR does not significantly associate with metabolic activity for (A) enalapril, (B) clopidogrel, and (C) methylphenidate. Horizontal lines indicate mean values of CES1 activity in each group.

Table 1

Relevant CES1 polymorphisms and their associations to drug response or gene expression.

SNP rs(number)	Gene	SNP Location/Base Change	Amino Acid Change/Location	Minor Allele Frequency	Function/Association
rs71647871	CES1	Chr16:55823658A>G	G143E	0.01	Loss of function, increased clopidogrel efficacy [10, 11]
rs3815583	CES1	Chr16:55833130G>T	5′UTR	0.287	Appetite reduction during methylphenidate treatment [12]
rs2244613	CES1	Chr16:55810697A>G	Intronic	0.333	Increased efficacy of dabigatran, incidence of sadness during methylphenidate treatment [13, 14]
rs3785161	CES1P1	chr16:55759783A>C	Promoter	0.1605	Increased promoter activity [8]

Table 2

CES1 SNPs introduced by the translocations from *CES1P1* short SNPs 1–5 (*CES1SVAR*) and long SNP 1–11 (*CES1VAR*)

A)

SNP	rs#	SNP	rs#	SNP	rs#
0	rs3815583	**4**	rs12149370	**8**	rs201577108(NS)
1	rs12149373	**5**	rs12149368	**9**	rs114788146(NS)
2	rs12149371	**6**	rs111604615(NS)	**10**	rs12149366(NS)
3	rs12149322	**7**	rs566557773(NS)	**11**	rs12149359

B)

	Major Allele	Minor Allele	Minor Allele Frequency
rs3815583 (Table 2A: SNP 0)	A	C	0.345
CES1VAR (Table 2A: SNPs 1–11)	CES1 sequence	CES1P1 5′ UTR, exon1 and intron 1 sequence	0.176
CES1SVAR (Table 2A: 1–5)	CES1 sequence	CES1P1 5′ UTR sequence	0.009
rs71647871 (G143E)	G	A	0.017

A) *CES1* 5′ region SNPs. *CES1VAR* and *CES1SVAR* are comprised of highly linked SNPs which match the sequence of *CES1P1*. SNPs 1–5 exist in the 5′ UTR, SNPs 6–10 in exon 1 and SNP 11 exists in intron 1. Non-synonymous variants are indicated below as “NS”.

B) Notable *CES1* allele frequencies.