Novel approach for *CES1* genotyping: integrating single nucleotide variants and structural variation

Ditte Bjerre¹, Henrik Berg Rasmussen^{*,1,2} & The INDICES Consortium³

¹Institute of Biological Psychiatry, Mental Health Centre Sct. Hans, Copenhagen University Hospital, DK-4000 Roskilde, Denmark

²Department of Science and Environment, Roskilde University, DK-4000 Roskilde, Denmark

³A list of the members of the consortium has been included in the Supplementary Materials accompanying this publication

* Author for correspondence: henrik.berg.rasmussen@regionh.dk

Aim: Development of a specific procedure for genotyping of *CES1A1* (*CES1*) and *CES1A2*, a hybrid of *CES1A1* and the pseudogene *CES1P1*. **Materials & methods:** The number of *CES1A1* and *CES1A2* copies and that of *CES1P1* were determined using real-time PCR. Long range PCRs followed by secondary PCRs allowed sequencing of single nucleotide variants in *CES1A1* and *CES1A2*. **Results & conclusion:** A procedure consisting of two main steps was developed. Its first main step, the copy number determination, informed about presence of *CES1A2*. This information enabled choice of PCR in the second main step, which selectively amplified *CES1A1* and, if present, also *CES1A2*, for subsequent sequencing. Examination of 501 DNA samples suggested that our procedure is specific with potential for personalization of drug treatments.

First draft submitted: 23 August 2016; Accepted for publication: 23 November 2017; Published online: 19 February 2018

Keywords: *CES1* • genotyping • long range PCR • single nucleotide variants • structural variation

CES1 is an $\alpha\beta$ -hydrolase fold enzyme belonging to the serine esterase superfamily [1]. This enzyme is a major contributor to the hydrolytic activity in the liver and metabolizes several commonly used therapeutic agents including methylphenidate [2], angiotensin-converting enzyme inhibitors [3,4], clopidogrel [5,6] and oseltamivir [7]. Moreover, *CES1* has an important role in the detoxification of pesticides [8,9]. *CES1* also appears to be involved in the hydrolysis of cholesteryl esters, triglycerides, and other endogenous compounds [10,11] suggesting physiological functions of this enzyme.

The region on chromosome 16 harboring *CES1* has a complex architecture, which primarily stems from gene duplication and the presence of a pseudogene, *CES1P1*, with a high degree of homology to *CES1* [12]. Additionally, there is another *CES1*-related pseudogene, *CES1P2*, more proximal on chromosome 16. The gene duplication event has led to the formation of a hybrid gene, which was designated *CES1A2*, whereas *CES1* commonly has been referred to as *CES1A1*. Both *CES1A1* and *CES1A2* have been redesignated as *CES1* in the recommended nomenclature for mammalian carboxylesterase gene families [13]. However, to distinguish between them we have maintained the designations, *CES1A1* and *CES1A2*, which we collectively refer to as *CES1A* in this publication.

The *CES1A2* region from the promoter to the initial part of intron 1 is derived from *CES1P1*, whereas the rest of this gene comes from *CES1A1*. Sequence differences in the promoter between *CES1A1* and *CES1A2* result in a significantly lower transcriptional level of *CES1A2* than *CES1A1* [14]. However, a promoter haplotype of *CES1A2* with two overlapping Sp1 binding sites and significantly increased transcriptional activity has been identified in one in five of this hybrid gene in a Japanese population [15].

Two subtypes of *CES1A1*, designated *CES1A1b* and *CES1A1c*, have been reported. Their exon 1 with or without the adjacent sequences were reported to be derived from *CES1P1* [14]. *CES1A1c*, the most common of the two subtypes, is also known as *CES1VAR*, whereas *CES1A1b* is identical with or closely related to *CES1VAR* [16].

Based on analysis of Caucasian, African-American and Japanese populations, two major structural *CES1* haplotypes have been reported [12]. The first of these is composed of *CES1P1* and *CES1A1*. In the second haplotype,

Figure 1. Haplotypes and structural variation in the chromosomal region harboring *CES1*. There are two major structural haplotypes that both carry *CES1*, also known as *CES1A1*. The first of these also contains *CES1P1*. In the second major structural haplotype this pseudogene has been partially replaced by *CES1A1* to form *CES1A2*. Based on the presence or absence of *CES1A1c* (*CES1VAR*), a variant of *CES1A1* with a *CES1P1* segment, these two major structural haplotypes can be subclassified, thus generating haplotypes A–D. *CES1A1* and *CES1P1* segments are shown in blue and magenta, respectively. Vertical bars and squares represent exons. Gray lines represent intergenic segments. Adapted from a previous publication [12].

CES1P1 has been replaced by *CES1A2*. Haplotypes with deletion of *CES1A1* have not been reported and appear to be rare. The two major structural haplotypes can be subclassified into haplotypes A–D based on the absence or presence of subtype *CES1A1c*. For graphical presentation of the four haplotypes, see Figure 1.

A variety of different approaches have been used in the genotyping of *CES1* including the 5' exonuclease-based method [17,18], allele-specific PCR [19,20], amplification followed by sequencing [21,22], chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometer [23] and the KASPar assay system [24]. Most of these approaches have not fully addressed the presence of *CES1A2* in some individuals and the high degree of sequence similarity between *CES1A1*, *CES1A2* and *CES1P1*. Accordingly, their specificity can be questioned. A more recent procedure based on long range amplification of a segment stretching from exon 1 to intron 6 of *CES1A1* and *CES1A2/CES1P1* identified the *CES1A1* subtype, *CES1A1c*, as *CES1A2* and was not able to distinguish *CES1A2* from *CES1P1*, thus excluding genotyping of *CES1A2* [25]. Collectively, this emphasizes that genotyping of *CES1* is not a trivial matter and may fail, in particular with standard procedures.

Recently, we reported a procedure based on long range PCR for amplification of a 12.5 kb fragment stretching from the promoter to intron 5 in *CES1A1* and *CES1A2*, respectively, followed by sequencing of amplified fragments [26]. We confirmed the specificity of this procedure including its ability to identify the *CES1A1* variants *CES1A1b* and *CES1A1c*, and that it distinguished *CES1A2* from *CES1P1*. Additionally, this procedure is capable of identifying the *CES1A2* promoter variant with the Sp1 sites.

Here we present a novel and integrated PCR-based genotyping procedure that involves determination of the combined copy number of *CES1A1* and *CES1A2* to guide downstream amplifications including amplification of the 12.5 kb fragment of *CES1A1* and *CES1A2*, respectively.

Materials & methods

DNA samples

Genomic DNAs from 400 individuals of Danish origin were included in the study. They had been isolated from samples of saliva or EDTA-stabilized blood. The samples of saliva were collected using the Oragene OG-250 DNA kit (DNA Genotek Inc., ON, Canada). We isolated genomic DNA from these samples using the prepIT L2P kit, produced by the manufacturer of the saliva collection kit. Isolation of genomic DNA from the blood samples was carried out using the Maxwell® 16 Instrument (Promega Corporation, WI, USA). We also obtained

45 and 56 samples of genomic DNA from Japanese and Han Chinese individuals, respectively, who had been enrolled in the 1000 Genomes Project (Coriell Institute for Medical Research, NJ, USA).

Determination of structural variation in *CES1A1* & *CES1P1*

The combined number of copies of *CES1A1* and *CES1A2*, that is, the number of *CES1A* copies, was determined. This was supplemented with copy number determination of *CES1P1*. The copy number determinations were carried out using commercially available reagents for duplex real-time PCR that quantitated the gene of interest and normalized it to an endogenous reference gene (Thermo Fisher Scientific, MA, USA). Two assays were used for determination of the combined copy number of *CES1A1* and *CES1A2*. They targeted regions in intron 11 (TaqMan[®] copy number assay Hs00139541_cn with location 55,836,763–55,867,075 on chromosome 16) and intron 8 (TaqMan[®] copy number assay Hs03948988_cn with location 55,836,763–55,867,075 on chromosome 16) of *CES1*, respectively. Since *CES1*, also referred to as *CES1A1*, and *CES1A2* are identical in these regions, both assays determine the combined number of copies of these two genes. For *CES1P1*, two assays, both targeting exon 6 of this pseudogene, were applied (TaqMan[®] copy number assay Hs01162042_cn and Hs01818263_cn with location 55,794,511–55,808,838 and 55,794,511–55,808,838 on chromosome 16, respectively). The gene encoding the ribonuclease P RNA component served as reference gene in the duplex real-time PCR (TaqMan[®] copy number reference assay RNase P, cat. #4403326 with location 20,811,565 on chromosome 14). The above chromosomal locations all refer to NCBI build 37. The assays were carried out according to the guidelines of the manufacturer of the reagents and calculations of the gene copy numbers were accomplished using CopyCaller v 1.0, a program developed by the reagent manufacturer. Information about the combined copy number of *CES1A1* and *CES1A2* was used to direct subsequent amplifications.

Primer design

We aligned the *CES1* reference sequence, NG_012057.1, from the *Homo sapiens* chromosome 16 assembly (GRCh38.p2 primary assembly: NC_000016.10), with the *CES1P1* and *CES1P2* sequences from the same assembly. Also included in the alignments were *CES1A2* (AB119998.1, complete coding sequence; and AB210090.1, promoter sequence). Regions that provided for a maximal difference between a desired sequence and the other sequences in the alignment were selected for design of primers. The specificity of the designed primers was assessed using Primer-BLAST [27] and the tool for *in silico* PCR tool in the UCSC genome browser [28].

Primers

Given that deletion of *CES1A1* is exceedingly rare, it is valid to assume for genomes with two *CES1A* copies that both of these represent *CES1A1* and that *CES1A2* is absent. Accordingly, *CES1A1* but not *CES1A2* were amplified from such genomes. A series of primer pairs was used for the amplifications of *CES1A1* and *CES1A2*. This included primer pairs for the selective amplification of fragments of 2.6 kb and 12.5 kb that stretched from the promoter to intron 1 and from the promoter to intron 5, respectively, of the two genes [26]. In order to amplify the region from intron 5 to the 3' untranslated region (3' UTR) of *CES1A1*, a primer pair amplifying a 19.2 kb fragment was designed. The locations of the primers for the amplification of the 2.6, 12.5 and 19.2 kb fragments are shown in Figure 2.

We also designed four primer pairs to amplify fragments that ranged from 375 to 659 bp in size and contained each of the exons 2–5 of *CES1A1* and *CES1A2*. Since these primer pairs did not differentiate between *CES1A1* and *CES1A2*, they were used in nested PCRs with the 12.5 kb amplicons of *CES1A1* and *CES1A2*, respectively, as template. Additionally, they were used for amplification using genomic DNA from individuals lacking *CES1A2* as template.

A set of primer pairs was exclusively used for nested PCR with the 19.2 kb *CES1A1* amplicon as template. This included a primer pair for the combined amplification of exon 6 and 7 of this gene and another that was used to amplify its exon 8. Sequences of the other primers for nested PCRs of *CES1A1*, that is, the primers for amplification of exon 9–14 derived from a previous study [22]. All primer sequences are listed in Supplementary Table 1.

Amplifications

The amplifications of the 2.6 and 12.5 kb fragments of *CES1A1* and *CES1A2* as well as the 19.2 kb *CES1A1* fragment were done using Herculase hotstart DNA polymerase (Agilent Technologies, Stratagene Products Division, CA, USA). This is an enzyme preparation, which consists of *Pfu* DNA polymerase, a small amount of *Taq2000*

Figure 2. Localization of master primers for PCR. Two primer pairs amplify a segment of 12.5 kb stretching from the promoter to intron 5 in *CES1A1* and *CES1A2*. Moreover, a 19.2 kb fragment of *CES1A1* is amplified that stretches from intron 5 to the 3' UTR. Two other primer pairs amplify a fragment of 2.6 kb stretching from the promoter to intron 1 in *CES1A1* and *CES1A2*, respectively. The 12.5 and 19.2 kb fragments are used as template in nested PCR to amplify and sequence a promoter segment and the coding regions. The 2.6 kb fragments of *CES1A1* and *CES1A2* are sequenced directly. *CES1A1* and *CES1P1* segments are shown in blue and red, respectively. The vertical bars represent exon 1–14. Note that *CES1A1* and *CES1A2* are shown in same transcriptional direction.

DNA polymerase, and a deoxyuridine triphosphatase (DUT). The latter removes 2'-deoxyuridine-5'-triphosphate formed by deamination of 2'-deoxycytidine-5'-triphosphate during PCR that would decrease strand elongation if incorporated. The amplifications were carried out in line with the guidelines of the manufacturer of the Herculanse hotstart DNA polymerase with minor modifications, which included incorporation of additional MgCl_2 in the reaction mixture for amplification of the 12.5 and 19.2 kb fragments and amplifications in volumes of 25 μl instead of 50 μl . We also used the Elongase[®] Enzyme Mix, which does not contain a DUT, for amplification of long fragments (Thermo Fisher Scientific, MA, USA). The amplifications using the Elongase[®] Enzyme Mix were done in accordance with the guidelines provided by the manufacturer of this enzyme mix.

Amplifications of fragments other than the 2.6, 12.5 and 19.2 kb fragments were performed using AmpliTaq Gold[®] DNA Polymerase (Thermo Fisher Scientific). For this purpose genomic DNA or the amplicons of 12.5 or 19.2 kb served as template after purification of the amplicons using the NucleoFast[®] 96 PCR purification Kit (Macherey–Nagel GmbH & Co KG, Düren, Germany).

Concentration of primers, template and annealing temperature were varied in order to determine the conditions that provided for the highest possible yield of the desired fragment without compromising the specificity or robustness of the amplifications. For composition of reaction mixtures and temperature cycling, see Supplementary Tables 3–11.

The primers for amplification using the proofreading DNA polymerase, were synthesized with a varying number of phosphorothioate (PTO) bonds to increase resistance toward the 3'-5' nuclease activity of the DNA polymerase that could otherwise compromise primer specificity by permitting elongation from mismatched primer–template complexes.

Electrophoretic analysis

Agarose gel electrophoresis was used to confirm that fragments of the expected sizes were amplified. We also used agarose gel electrophoresis as a mean to determine the reactions conditions that produced high yields of the fragments of the expected sizes without amplification of fragments of other sizes, that is, undesired fragments.

Figure 3. Flowchart of *CES1* typing approach. The combined copy number of *CES1A1* and *CES1A2* is used to inform about the presence or absence of *CES1A2* and select downstream amplifications. Samples with a copy number of two are not expected to contain *CES1A2*. A 12.5 kb fragment stretching from the promoter to intron 5 and a 19.2 kb fragment stretching from intron 5 to the 3' UTR of *CES1A1* are amplified from such samples. The 12.5 kb *CES1A1* fragment and the corresponding *CES1A2* fragment are amplified from genomes with a copy number of three or four.

Sanger sequencing

Amplified products were purified using the QIAquick 96 PCR Purification Kit (Qiagen NV, Venlo, The Netherlands) and subjected to sequencing by the Sanger dideoxy chain termination method. For this purpose the amplification primers were used as sequencing primers. Primers specifically dedicated to sequencing were also used; see Supplementary Table 2.

Analysis of sequences

Sequencing chromatograms were visually inspected for traces of mixed sequences suggesting presence of two or more templates and lack of specificity. Subsequently, sequences were aligned with the *CES1* reference sequence, NG_012057.1, from the *Homo sapiens* chromosome 16 assembly (GRCh38.p2 primary assembly: NC_000016.10), the *CES1P1* sequence from the same assembly and the *CES1A2* sequence with accession number AB119998.1. Also included in the alignments were the sequences with the accession numbers AB210087.1 (*CES1A1a*), AB210088.1 (*CES1A1b*), AB210089.1 (*CES1A1c*) and AB210090.1 (*CES1A2*). The promoter variant of *CES1A2* with increased transcriptional activity was identified based on sequence data from a previous study [15].

Results

We developed an integrated approach in which information from the copy number determinations of *CES1A* and *CES1P1* was used to select the subsequent amplifications aimed at identification and genotyping of single nucleotide variants (SNVs) in *CES1A1* and *CES1A2* by sequencing. An overview of the procedure is shown in Figure 3. Our approach consisted of two main steps. The first of these steps was the copy number determinations in which the *CES1P1* copy number served to confirm the *CES1A* copy number since the total number of copies of *CES1A1*, *CES1A2* and *CES1P1* adds to four [12]. The second main step consisted of the amplification of different PCR fragments depending upon the absence or presence of *CES1A2*.

For genomes with a *CES1A* copy number of two, that is, genomes lacking *CES1A2*, we amplified two *CES1A1* fragments of 12.5 and a 19.2 kb that stretched from the promoter to intron 5 and from intron 5 to the 3' UTR of the gene, respectively. Due to sequence similarity between *CES1A1* and *CES1A2*, the reaction for production of the

Figure 4. Agarose gel electrophoresis of 2.6 kb fragments of *CES1A1* and *CES1A2* amplified using genomes with a *CES1A* copy number of two or three as template. Lane 1: sample with a *CES1A* copy number of two subjected to amplification using the *CES1A1* primer pair. Lane 2: sample with a *CES1A* copy number of three subjected to amplification using the *CES1A1* primer pair. Lane 3: sample with a *CES1A* copy number of two subjected to amplification using the *CES1A2* primer pair. Lane 4: sample with a *CES1A* copy number of three subjected to amplification using the *CES1A2* primer pair. Note that the sample with a *CES1A* copy number of two does not support amplification using the *CES1A2* primer pair. The *CES1A* copy number refers to the combined number of copies of *CES1A1* and *CES1A2*. To the left fragments of a DNA molecular weight marker in base pairs. For primers, see Supplementary Table 1.

19.2 kb *CES1A1* fragment also amplified the corresponding *CES1A2* fragment, thus not being applicable to the analysis of genomes with a combined *CES1A1* and *CES1A2* copy number of three or four. Hence, amplification using genomes with such copy numbers as template was restricted to the production of the 12.5 kb fragments of *CES1A1* and *CES1A2*.

The fragments obtained by the long PCRs were used as templates in nested PCRs for secondary amplification of a promoter fragment and the coding regions. Using as template the 12.5 kb fragment of *CES1A1* and *CES1A2*, respectively, we amplified the 2.6 kb fragment containing a promoter segment and exon 1, which allowed for the identification of *CES1A1c*, and fragments with sizes that ranged from 375 to 659 bp and contained exon 2–5. The 19.2 kb *CES1A1* fragment served as template for nested PCRs to amplify exon 6–14. Accordingly, all *CES1A1* exons and a portion of the promoter were amplified from individuals lacking *CES1A2*.

We replaced the 12.5 kb amplicon with genomic DNA as template in the reaction mixtures for amplification of the 2.6 kb fragment containing the promoter and exon 1 of *CES1A1* and *CES1A2*. Moreover, we amplified the *CES1A1* exons 2–5 in separate reaction using genomic DNA from individuals lacking *CES1A2* as template. Since the primer pairs for amplification of these exons did not differentiate between *CES1A1* and *CES1A2*, they were not applicable for amplification of samples harboring *CES1A2* with genomic DNA as template.

Primer sequences, amplicons, optimized reaction mixtures and optimized cycling conditions are listed in Supplementary Tables 1 & 3–11.

The use of PTO-modified nucleotides successfully prevented unspecific amplification in the reactions designed for production of the 2.6, 12.5 kb and 19.2 kb amplicons. Up to four PTO-modified nucleotides in the 3' end of the primers were required to prevent unspecific amplification effectively.

The enzyme preparation containing DUT produced significantly larger amounts of the 2.6, 12.5 and 19.2 kb fragments as compared with the preparation without it. Hence, the former of these enzyme mixtures was preferred.

The specificity of the amplifications was assessed by different means. This included analysis of the *CES1A1* primers by *in silico* PCR, which predicted amplification of the desired fragments from this gene. The *in silico* PCR did not predict product outcomes for the primer pairs designed to amplify the 2.6 and 12.5 kb fragments of *CES1A2*. This was expected as *CES1A2* is not included in the current or earlier versions of the build of the human genome used by the *in silico* PCR programs for product outcome prediction. The gel electrophoretic analysis revealed that the amplifications all produced fragments of the expected sizes. Assessment of the amplification specificity by inspection of the sequencing chromatograms did not reveal traces of mixed sequences reflecting co-amplification of undesired fragments such as *CES1P1* fragments. Moreover, the sequences of the amplified fragments aligned to the genes, which they had been designed to amplify. We also found that samples with a combined number of *CES1A1* and *CES1A2* copies of two did not support amplification of the 2.6 and 12.5 kb fragments using the *CES1A2* primers, while these primers amplified fragments of the expected sizes from genomes with a total of three or four of these copies (Figure 4).

We used the newly developed approach for examination of 501 samples and found that the copy numbers of *CES1A* and *CES1P1* were unambiguously determined in 495 out of these samples. The remaining samples were discarded due to poor DNA quality. The *CES1A* copy number was found to range from two to four while that of *CES1P1* ranged from zero to two with the total number of copies of *CES1A1*, *CES1A2* and *CES1P1*

Figure 5. Sanger sequencing chromatogram of a segment of *CES1A1* exon 4 containing p.Gly143Glu (rs121912777). Note a GG homozygote (upper lane), a GA heterozygote (middle lane) and an AA homozygote (lower lane) at the position denoted with an R. The two peaks of the heterozygote at this position are of almost half the heights of the homozygote peaks at the same position.

amounting to four. All of the examined samples supported amplification with the primers designed to amplify the 2.6 and 12.5 kb fragments of *CES1A1*, which is in line with the observation that deletion of *CES1A1* is rare [12]. On average about 97% of the amplicons produced readable sequences. We identified several previously reported *CES1A1* variations such as p.Gly143Glu (rs121912777), which has been associated with significantly decreased catalytic activity [18], subtype *CES1A1c* [14] and the *CES1A2* variant with the two overlapping Sp1 sites conferring increased transcriptional activity and potentially rapid drug metabolism [15]. For Sanger sequencing chromatogram of *CES1A1* p.Gly143Glu, see Figure 5.

We compared findings from the present study with those obtained by a previous study, which examined the diversity of the gene encoding CES1 in a European population and reported a total of 16 SNVs in the region from the promoter to the 3' UTR of this gene without distinguishing between *CES1A1* and *CES1A2* [21]. We detected a significantly higher number of SNVs, namely, 46 and 21 in *CES1A1* and *CES1A2*, respectively, in our Danish population. This included 15 SNVs in the coding regions of *CES1A1* and *CES1A2* of which five were detected in *CES1A1* exon 1 at frequencies of about 0.20, namely, rs111604615, rs566557773, rs201577108, rs114788146 and rs12149366. None of these five SNVs, which all are a part of *CES1A1c* (*CES1VAR*) [16], were detected in the previous study [21]. We are not aware of other studies of the diversity of *CES1A1* in European populations. Comparison of our findings with those of the 1000 Genomes Project [29] also revealed major differences; see Table 1. Notably, the 1000 Genomes Project reported minor allele frequencies (MAFs) in the range from 0.004 to 0.080 for the SNVs that define *CES1A1c* in exon1 in European and east Asian populations, whereas we found MAFs of up to 0.29 for some of these SNVs in the same populations.

Discussion

We report the development of a novel approach with a low drop-out rate for genotyping of *CES1A1* (*CES1*). This development is based on copy number determination to deduce the structural haplotypes and guide subsequent amplification of fragments of *CES1A1* and *CES1A2*. Sequencing of these fragments permits the identification of SNVs and other small-scale variations. Our approach distinguishes itself by the application of copy number determination and its ability to selectively amplify fragments of *CES1A1* and *CES1A2*. For example, it allowed the identification of the *CES1A2* variant with the two overlapping Sp1 sites conferring increased transcriptional activity and potentially rapid drug metabolism.

The number of copies of *CES1A1*, *CES1A2* and *CES1P1* adds to four with two of these being *CES1A1* since deletion of *CES1A1* appears to be extremely rare [12]. This 'four rule' allowed us to use the copy number of *CES1P1* to validate the results from the copy number determination of *CES1A1* and *CES1A2*. This is useful as the number of copies of a gene cannot always be determined unambiguously due to inferior DNA quality or residual impurities. For example, it is occasionally unclear whether a gene is present in two or three copies reflecting that a copy number

Table 1. Single nucleotide variants in the 5' end of *CES1A1*.

Location [†]	dbSNP reference	Minor allele frequency			
		Our procedure		1000 Genomes Project [‡]	
		Europeans (Danes)	Chinese & Japanese	Europeans	East Asians
5' UTR	rs12149373	0.209	0.271	0.164	0.221
5' UTR	rs12149371	0.216	0.284	0.176	0.218
5' UTR	rs12149322	0.210	0.280	0.082	0.110
5' UTR	rs12149370	0.212	0.280	0.082	0.110
5' UTR	rs12149368	0.212	0.273	0.053	0.072
Exon 1	rs111604615	0.210	0.270	0.004	0.011
Exon 1	rs566557773	0.211	0.270	0.025	0.054
Exon 1	rs201577108	0.211	0.270	0.025	0.054
Exon 1	rs114788146	0.211	0.270	0.035	0.056
Exon 1	rs12149366	0.211	0.276	0.067	0.080
Intron 1	rs12149359	0.208	0.273	0.125	0.156

[†]The single nucleotide variants in the 5' UTR, exon 1 and intron 1 that define *CES1A1c*.

[‡]Data were extracted from dbSNP [30].

of three differs from a copy number of two by a factor of only 1.5. By contrast, samples with one copy of a gene can better be discriminated from samples having two copies, since a copy number of one differs from a copy number of two by a factor of two. According to the 'four rule', *CES1P1* copy numbers of one and two correspond to a combined *CES1A1* and *CES1A2* copy number of three and two, respectively.

Challenges to the selective amplification of the fragments of *CES1A1* and *CES1A2* are the high sequence identity between these two genes and their sequence identity with *CES1P1*. Another challenge is the requirement for a proofreading DNA polymerase in the amplification of the long fragments since such a polymerase repairs mismatches between a primer and a template, thus potentially permitting strand elongation and amplification with concomitant loss of specificity. This was overcome by the use of primers modified with PTO to provide nuclease resistance. Additionally, the enzyme mixture preferred for amplification of the large fragments contained a DUT, which removes 2'-deoxyuridine-5'-triphosphate during PCR, thus promoting strand elongation [31]. This enzyme mixture greatly enhanced the amplification of the 2.6, 12.5 and 19.2 kb fragments compared with a preparation without DUT and enabled a robust amplification of these fragments.

Since the majority of hepatic CES1 is produced by *CES1A1* [12,14], variation in this gene would be expected to have a significantly larger impact on the enzyme activity than variation in *CES1A2*. Hence, it is critical to distinguish between *CES1A1* and *CES1A2* in the assessment of the functional impact of a *CES1A* variant. This requires a genotyping approach with high resolution and specificity, which was provided by our approach.

Several of the previously reported procedures for genotyping of *CES1A1* and *CES1A2* appear not fully to have taken the complex pattern of variants into consideration. This included a study that reported a variety of SNVs in *CES1* based on amplification followed by sequencing of amplicons [21]. Using *in silico* PCR we recently found that all of the primers from this study amplified *CES1A1* as well as *CES1A2*, thus producing mixed amplicons [34]. In line with this, we found major differences in the number of SNVs and their MAFs between our study and the previous study [21], although some of these differences may be explained by the amplification and examination of fragments that were overlapping but not identical and differences in the number of subjects in the two studies. Another study examined variation in *CES1* and *CES1P1* also without distinguishing between *CES1A1* and *CES1A2*, thus potentially leading to inaccurate variant detection [22]. Moreover, the design and location of our primers represents an advantage compared with the previous procedure based on long PCR [25], which did not distinguish *CES1A2* from *CES1P1* and also lacked the ability to distinguish *CES1A2* from *CES1A1c*.

The specificity of the amplifications in our method was assessed by *in silico* PCR, agarose gel electrophoresis and Sanger sequencing. We also examined whether genomes without *CES1A2* supported amplification using *CES1A2* primer pairs. The observations achieved by these means supported each other and indicated that our amplification reactions were specific with the ability to selectively amplify *CES1A1* and *CES1A2*, respectively, without co-amplification of *CES1*-related pseudogenes or other undesired fragments.

A limitation to our approach is that it is not able to genotype and identify nucleotide variants from intron 5 and downstream in *CES1A1* and *CES1A2* in individuals carrying both of these genes. Hence, potentially damaging SNVs may not be detected in some individuals. The inability to detect variants in the downstream regions of *CES1A1* and *CES1A2* in individuals carrying both of these genes reflects their sequence similarity.

The complexity in the chromosomal region harboring the *CES1* locus with the presence of a hybrid of *CES1P1* and *CES1* in some individuals has a parallel in the gene encoding *CYP2D6*. That is, hybrids of *CYP2D6* and the pseudogene *CYP2D7* have been reported to complicate the genotyping of *CYP2D6*, thus requiring specifically designed assays [32]. Additionally, a variation in pseudogene *CYP2D7* has been shown to permit binding of a primer for identification of *CYP2D6* alleles resulting in false positive calls for these alleles [33]. Whether this also is a problem in the typing of *CES1* is not yet known.

An accurate procedure for detection of variants in *CES1* may be valuable as a tool to assist in predicting efficacy and risk of adverse reactions of therapeutic agents metabolized by *CES1* and reveal individual predisposition to illicit drugs and toxicants that are inactivated by this enzyme. It may also be helpful to provide new insights into the genetics of lipid metabolism and obesity.

We developed such procedure, which takes the structural variation of *CES1* into consideration and integrates determination of structural variation with genotyping of SNVs and other small-scale variants. This permits correct annotation of SNVs in *CES1A1* and *CES1A2* and enables identification of the subtypes of *CES1A1* and the *CES1A2* variant associated with increased transcriptional activity. We propose our procedure to become a standard in the genotyping of *CES1*.

Conclusion

We developed a procedure for genotyping of *CES1A1* (*CES1*) and *CES1A2*, a hybrid of *CES1A1* and the *CES1*-related pseudogene *CES1P1*. This procedure is based on copy number determination to inform about the presence or absence of *CES1A2* in an individual followed by amplification of fragments for genotyping of small-scale variants by Sanger sequencing. Our procedure appeared to be highly specific and distinguished itself from most previous procedures by taking into account the presence of *CES1A2* in some individuals and the large degree of sequence homology between *CES1A1*, *CES1A2* and *CES1P1*.

Future perspective

Development of methods for genotyping of *CES1A1* and *CES1A2* with high specificity permits a reappraisal of the diversity of these genes and inclusion of more accurate data in genetic variant databases. Future studies are also believed to address the lack in the ability to genotype a large portion of the 3' ends of *CES1A1* and *CES1A2* in individuals carrying both of these genes. The development and validation of novel analyses for genotyping of *CES1A1* and *CES1A2* would be facilitated by detailed characterization of these genes in selected samples that could be included in genetic resources repositories as reference samples.

Note

After initiation of the present method development the single nucleotide substitution rs74518148 was reported. The variant nucleotide at this polymorphic site creates a mismatch with our forward primer for amplification of the 2.6 and 12.5 kb fragments of *CES1A1* (*CES1*). Since this mismatch is located in the 5' end of the primer and the minor allele of rs74518148 has a low frequency, we do not anticipate it to affect the amplification outcome. However, we now synthesize the primer concerned with mixed bases (underlined) to achieve complete match with both alleles at rs74518148: 5'- ACTAYGGGGGGGACGGAGTTCA-3'.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at:

<https://www.futuremedicine.com/doi/suppl/10.2217/pgs-2016-0145>

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Background

- *CES1A1* (*CES1*) is a pharmacogenetic candidate gene and encodes a hydrolase, which is implicated in the metabolism of commonly used drugs.
- The chromosomal region harboring *CES1A1* is complex due to the existence of a pseudogene, *CES1P1*, and a structural variant known as *CES1A2*, which is a hybrid of *CES1A1* and *CES1P1*.
- Several previous genotyping methods have not been able to distinguish between *CES1A1*, *CES1A2* and *CES1P1*.
- We aimed at developing a specific method for the genotyping of *CES1A1* and *CES1A2*.

Materials & methods

- Real-time PCR served to detect the presence or absence of *CES1A2* in a sample.
- Primer pairs amplifying 2.6 and 12.5 kb fragments of *CES1A1* and *CES1A2* in addition to a primer pair producing a 19.6 kb *CES1A1* amplicon, were designed in regions of these genes that enabled highest possible degree of specificity.
- Amplification primers containing varying numbers of phosphorothioate bonds in their 3' ends and different DNA polymerases were tested.
- Amplicons were sequenced by the Sanger method.
- Samples of DNA from 400 individuals residing in Denmark and 101 individuals from the 1000 Genomes Project were used for development of the new procedure.

Results

- We developed a procedure for genotyping of *CES1A1* and *CES1A2* consisting of two main steps with copy number determination by real-time PCR as the first main step serving to guide amplification of fragments in the second main step of the procedure.
- In the main second step fragments were amplified for Sanger sequencing using a proofreading DNA polymerase and primers containing phosphorothioate bonds, which prevented loss in specificity as a result of repair of mismatches between the primers and undesired targets.
- Our procedure appeared to be specific with the ability to distinguish *CES1A1* from *CES1A2* but is unable to genotype variants in the downstream half of *CES1A1* and *CES1A2* in individuals carrying both of these genes.

Discussion

- Structural variation complicates the genotyping of single nucleotide variations in *CES1A1* and *CES1A2* markedly.
- Our procedure opens up for a more accurate genotyping of *CES1A1* and *CES1A2* and may be helpful in the individualization of treatments with drugs metabolized by *CES1*.

References

1. Cygler M, Schrag JD, Sussman JL *et al*. Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. *Protein Sci.* 2(3), 366–382 (1993).
2. Sun Z, Murry DJ, Sanghani SP *et al*. Methylphenidate is stereoselectively hydrolyzed by human carboxylesterase CES1A1. *J. Pharmacol. Exp. Ther.* 310(2), 469–476 (2004).
3. Takai S, Matsuda A, Usami Y *et al*. Hydrolytic profile for ester- or amide-linkage by carboxylesterases pI 5.3 and 4.5 from human liver. *Biol. Pharm. Bull.* 20(8), 869–873 (1997).
4. Zhu H-J, Appel DI, Johnson JA, Chavin KD, Markowitz JS. Role of carboxylesterase 1 and impact of natural genetic variants on the hydrolysis of trandolapril. *Biochem. Pharmacol.* 77(7), 1266–1272 (2009).
5. Hagihara K, Kazui M, Kurihara A *et al*. A possible mechanism for the differences in efficiency and variability of active metabolite formation from thienopyridine antiplatelet agents, prasugrel and clopidogrel. *Drug Metab. Dispos.* 37(11), 2145–2152 (2009).
6. Zhu H-J, Wang X, Gawronski BE, Brinda BJ, Angiolillo DJ, Markowitz JS. Carboxylesterase 1 as a determinant of clopidogrel metabolism and activation. *J. Pharmacol. Exp. Ther.* 344(3), 665–672 (2013).
7. Shi D, Yang J, Yang D *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.* 319(3), 1477–1484 (2006).
8. Ross MK, Borazjani A, Edwards CC, Potter PM. Hydrolytic metabolism of pyrethroids by human and other mammalian carboxylesterases. *Biochem. Pharmacol.* 71(5), 657–669 (2006).
9. Ross MK, Crow JA. Human carboxylesterases and their role in xenobiotic and endobiotic metabolism. *J. Biochem. Mol. Toxicol.* 21(4), 187–196 (2007).
10. Quiroga AD, Lehner R. Role of endoplasmic reticulum neutral lipid hydrolases. *Trends Endocrinol. Metab.* 22(6), 218–225 (2011).
11. Xie S, Borazjani A, Hatfield MJ, Edwards CC, Potter PM, Ross MK. Inactivation of lipid glyceryl ester metabolism in human THP1 monocytes/macrophages by activated organophosphorus insecticides: role of carboxylesterases 1 and 2. *Chem. Res. Toxicol.* 23(12), 1890–1904 (2010).

12. Fukami T, Nakajima M, Maruichi T *et al.* Structure and characterization of human carboxylesterase 1A1, 1A2, and 1A3 genes. *Pharmacogenet. Genomics* 18(10), 911–920 (2008).
13. Holmes RS, Wright MW, Laulederkind SJF *et al.* Recommended nomenclature for five mammalian carboxylesterase gene families: human, mouse, and rat genes and proteins. *Mamm. Genome*. 21(9–10), 427–441 (2010).
14. 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* 17(1), 1–10 (2007).
15. Yoshimura M, Kimura T, Ishii M *et al.* Functional polymorphisms in carboxylesterase 1A2 (*CES1A2*) gene involves specific protein 1 (Sp1) binding sites. *Biochem. Biophys. Res. Commun.* 369(3), 939–942 (2008).
16. Sanford JC, Wang X, Shi J *et al.* Regulatory effects of genomic translocations at the human carboxylesterase-1 (*CES1*) gene locus. *Pharmacogenet. Genomics* 26(5), 197–207 (2016).
17. Nemoda Z, Angyal N, Tarnok Z, Gadoros J, Sasvari-Szekely M. Carboxylesterase 1 gene polymorphism and methylphenidate response in ADHD. *Neuropharmacology* 57(7–8), 731–733 (2009).
18. Zhu H-J, Patrick KS, Yuan H-J *et al.* Two *CES1* gene mutations lead to dysfunctional carboxylesterase 1 activity in man: clinical significance and molecular basis. *Am. J. Hum. Genet.* 82(6), 1241–1248 (2008).
19. Ding X-L, Deng Y-L, Zhang J, Miao L-Y. Mutation-sensitive molecular switch method to detect *CES1A2* mutation in the Chinese Han and Yao populations. *Genet. Test Mol. Biomarkers* 15(9), 659–662 (2011).
20. Soria NW, Belas A, Galván C *et al.* A simple allele-specific polymerase chain reaction method to detect the Gly143Glu polymorphism in the human carboxylesterase 1 gene: importance of genotyping for pharmacogenetic treatment. *Genet. Test. Mol. Biomarkers* 14(6), 749–751 (2010).
21. Marsh S, Xiao M, Yu J *et al.* Pharmacogenomic assessment of carboxylesterases 1 and 2. *Genomics* 84(4), 661–668 (2004).
22. Yamada S, Richardson K, Tang M *et al.* Genetic variation in carboxylesterase genes and susceptibility to isoniazid-induced hepatotoxicity. *Pharmacogenomics J.* 10(6), 524–536 (2010).
23. Zou J-J, Chen S-L, Fan H-W, Tan J, He B-S, Xie H-G. *CES1A* -816C as a genetic marker to predict greater platelet clopidogrel response in patients with percutaneous coronary intervention. *J. Cardiovasc. Pharmacol.* 63(2), 178–183 (2014).
24. Johnson KA, Barry E, Lambert D *et al.* Methylphenidate side effect profile is influenced by genetic variation in the attention-deficit/hyperactivity disorder-associated *CES1* gene. *J Child Adolesc. Psychopharmacol.* 23(10), 655–664 (2013).
25. Zhu H-J, Brinda B, Froehlich TE, Markowitz JS. A discriminative analytical method for detection of *CES1A1* and *CES1A2/CES1A3* genetic variants. *Pharmacogenet. Genomics* 22(3), 215–218 (2012).
26. Bjerre D, Rasmussen HB; INDICES Consortium. Novel procedure with improved resolution and specificity for amplification and differentiation of variants of the gene encoding carboxylesterase 1. *Pharmacogenet. Genomics* 27(4), 155–158 (2017).
27. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13, 134 (2012).
28. Speir ML, Zweig AS, Rosenbloom KR *et al.* The UCSC genome browser database: 2016 update. *Nucleic Acids Res.* 44(D1), D717–D725 (2015).
29. 1000 Genomes Project Consortium; Auton A, Brooks LD *et al.* A global reference for human genetic variation. *Nature* 526(7571), 68–74 (2015).
30. Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. (dbSNP Build ID: 151).
31. Hogrefe HH, Hansen CJ, Scott BR, Nielson KB. Archaeal dUTPase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation. *Proc. Natl Acad. Sci. USA* 99(2), 596–601 (2002).
32. Gaedigk A, Fuhr U, Johnson C, Bérard LA, Bradford D, Leeder JS. *CYP2D7-2D6* hybrid tandems: identification of novel *CYP2D6* duplication arrangements and implications for phenotype prediction. *Pharmacogenomics* 11(1), 43–53 (2010).
33. Riffel AK, Dehghani M, Hartshorne T *et al.* *CYP2D7* sequence variation interferes with TaqMan *CYP2D6*15* and **35* genotyping. *Front. Pharmacol.* 6, 312 (2015).
34. Rasmussen HB, Madsen MB; INDICES Consortium. Carboxylesterase 1 genes – systematical review and evaluation of existing genotyping procedures. *Drug Metab. Pers. Ther.* (2018) (Epub ahead of print).