Diversity of common alternative splicing variants of human *Cytochrome P450 1A1* and their association to carcinogenesis

MARIO BAUER[1], OLF HERBARTH[1], SUSANNE RUDZOK[1], EIKE SCHMÜCKING[1], ANDREA MÜLLER[1], GABRIELA AUST[2] and CAROLIN GRÄBSCH[1]

[1]Dept. of Human Exposure Research and Epidemiology, Helmholtz Centre for Environmental Research - UFZ, Permoserstr. 15, D-04318 Leipzig; [2]Research Laboratories, Department of Surgery, University of Leipzig, Leipzig, Germany

Received January 9, 2007; Accepted March 19, 2007

Abstract. Cytochrome P450 1A1 (CYP1A1) belongs to the enzymes of biotransformation of phase I. CYP1A1 performs the catalytic activation of exogenous and endogenous substrates to more carcinogenic metabolites. Overexpression of the wild-type and a recently described splice variant (*CYP1A1v*, ovarian cancer) are attributed to neoplastic transformation. Here we describe novel *CYP1A1* splicing variants commonly and frequently transcribed in leucocytes of healthy volunteers, separated from variants exclusively expressed in tumour cell lines. Interestingly, all the novel splicing variants in leukocytes are generated by employing of two nested splice site pairs, one outer canonical and one inner non-canonical splice site pair, within the exon 2 of the human *CYP1A1*. In general, the frequent presence of common splicing variants in healthy volunteers has to be consider as a physiological feature of human CYP1A1 transcription process, rather than a signature of carcinogenesis.

Introduction

Among the various forms of CYPs determined so far, CYP1A1 and CYP1B1 have been shown to be the most important human CYP enzymes in metabolizing procarcinogens, such as PAHs and aromatic amines, into active species forming DNA adducts (1-4).

The constitutive and inducible expression of CYP1A1 and CYP1B1 are considered to be important determinants of carcinogenesis. Induction of CYP1A1 would be expected to increase the activation of procarcinogens to DNA-reactive metabolites, leading to increased tumour formation. Inducible CYP1A1 activity is ubiquitous, located in virtually every tissue of the body including endothelial cells of blood vessels, leukocytes, epithelial cells of the skin and gastrointestinal tract, fetus, and embryo (5).

Overexpression is frequent after exposure toward different environmental factors such as cigarette smoke or local pollutants (6) and subsequently in there initiating tumour tissue. The expression level and microsomal CYP1A1 protein level are significantly higher among current smokers compared to ex-smokers and never-smokers in the lung (7,8). A similar high increased CYP1A1 expression is observed in different human oral cell phenotypes after cigarette smoke condensate (CSC) exposure. For other cancers such as gastric cancer the frequencies of CYP1A1 expression are significantly different between non-cancerous and premalignant groups, suggesting that CYP1A1 is expressed at relatively early stage of gastrocarcinogenesis and exerts its effects throughout the stepwise oncogenic processes (9). However, the exact relationship between CYP1A1 overexpression and chemically induced carcinogenesis remains to be established (10).

Contrary to the expected association between CYP1A1 overexpression and carcinogenesis, there is little evidence from other human epidemiological studies or experiments on animals that P450 induction enhance the incidence or multiplicity of tumours caused by known chemical carcinogens (11). Although smoking is a major risk factor for the common esophageal cancer, DNA adducts were associated with upregulation of CYP1B1, but not CYP1A1, mRNA and protein in human esophageal HET-1A cells (12). For breast cancer a low frequently expressed CYP1A1 (13), similar to low *CYP1A1* mRNA expression levels in neoplastic and non-neoplastic tissue (14-16), has been described. Other studies report an unexpected not overexpressed CYP1A1 level in hepatocellular carcinoma (HCC) and squamous cell carcinoma of the head and neck (SCCHN) (17,18).

An overexpressed CYP1A1 has been assumed to play a significant role in DNA adduct formation which support a higher risk for cancer initiation. CYP1A1 is mainly localized in the endoplasmic reticulum. In contrast, a recently described, enzymatic active, spliced variant of CYP1A1 (*CYP1A1v*, GenBank accession no. AY310359) exhibits a nucleus and mitochondria restricted subcellular distribution. The nuclear distribution evidently contributes to ovarian cancer initiation and progression additionally to an overexpressed CYP1A1 (19).

Correspondence to: Dr Mario Bauer, Dept. of Human Exposure Research and Epidemiology, Helmholtz Centre for Environmental Research - UFZ, Permoserstr. 15, D-04318 Leipzig, Germany
E-mail: mario.bauer@ufz.de

Key words: Cytochrome P450 1A1, alternative splicing, human leukocytes, human epithelial tumour cell line

Detecting full-length mRNA transcripts of participants of epidemiological studies we describe the common presence of splicing variants of CYP1A1 in healthy volunteers including the enzymatic active variant *CYP1A1v*. We argue that their common presence should be consider as a physiological signature rather than a signature of carcinogenesis. An overview of frequently present multiple splicing variants of different human cancer cell lines in comparison to blood leucocytes of healthy volunteers is given.

Materials and methods

Human probes. The non-adherent growing human small-cell lung cancer (SCLC) cell lines NCI-H82 (20), NCI-H69 (21), the non-SCLC cell line MR65, the adherent growing non-SCLC cell lines A549 (22), Colo-699, LCLC-103H (23), Oka-C-1 (24), EPLC-272H (25), KNS-62 (26), NCI-H322 (27), NCI-H358 (28), BEN (29), the hepatocyte cell line HepG2 (30), the cervix carcinoma cell line HBL-100 and the ovary adenocarcinoma cell lines EFO-21 (serous cystadenocarcinoma), -27 (mucinous papillary adenocarcinoma) (31) were obtained from European Collection of Cell Cultures and German Collection of Microorganisms and Cell Cultures.

The human hepatocyte cell lines IHH1 (immortalized human hepatocyte no. 1), HH1 (human hepatoma no. 1), IHFL1 (immortalized human fetal liver cells no. 1) and IHI1 (immortalized human islet cells no. 1) were kindly provided by Jan Hengstler from the University of Leipzig. The phenotype of hepatocytes was controlled by measurement of glucose, lactate and urea in the supernatant and the CYP3A4 and GST activity of the cell homogenate.

Lung and cervix carcinoma cell lines were kept in RPMI-1640 (PAN Biotech, Aidenbach, Germany) supplemented with 10% fetal calf serum (FCS, Biochrom AG, Berlin, Germany). Hepatocyte cell lines were held in William's Medium E/10% FCS. The culture medium of IHI1 cell line was additionally supplemented with 20 ng/ml basic fibroblast growth factor (bFGF, PAN) and 20 ng/ml epidermal growth factor (EGF, PAN). Ovary adenocarcinoma cell lines were kept in DMEM (PAN) supplemented with 10% FCS.

Human venous blood samples were taken from healthy volunteers in accordance with protocols of the local Committee of Medical Ethics. All participants gave their written consent. The mononuclear fraction (MNC) was separated by separation media (PAA, Cölbe, Germany), according to the manufacturer's instruction. The granulocyte fraction was prepared after red blood cell lysis of the separation sediment. The MNC and granulocytes were held thereafter in RPMI-1640 media/10% autologous serum for 24 h.

Isolation and amplification of nucleic acid. Cellular RNA was isolated from human cell lines as well as from blood samples with the peqGold RNAPure™ Isolation protocol (Peqlab, Erlangen, Germany) according to the manufacturer's instruction. After a DNAse (Roche, Mannheim, Germany) digestion step, 5 μg of total RNA was taken to synthesize cDNA using the RevertAid H Minus First strand cDNA Synthesis Kit (Fermentas GmbH, St. Leon-Rot, Germany). The primer *CYP1A1*_for1 5'-TGA TCC CAG GCT CCA AGA GTC CAC-3', *CYP1A1*_for2 5'-TCC CAG CTC AGC TCA GTA CCT CAG-3', *CYP1A1*_for3 5'-CAG TAC CTC AGC CAC CTC CAA GAT-3', *CYP1A1*_rev1 5'-ATC AGG GGT GAG AAA CCG TTC AGG-3' and *CYP1A1*_rev2 5'-GAC AGC TGG ACA TTG GCG TTC TC-3' (MWG Biotech, Ebersberg, Germany) were used to perform a nested PCR approach detecting transcripts from the first to the last exon. Transcript lengths of 1384 (*CYP1A1_for1/-rev1*), 1357 (*-for2/-rev1*), 1344 (*-for3/-rev1*), 958 (*-for2/-rev2*) and 945 (*-for3/-rev2*) bp are expected. The primers were designed without mismatches to common single nucleotide polymorphisms to exclude false negative results. The PCR was performed at 94˚C for 5 min, then 38 cycles at 94˚C for 30 sec, at 60˚C for 30 sec, at 72˚C for 90 sec, and extension at 72˚C for 10 min. The PCR mixture contained 1.25 mM $MgCl_2$, 200 μM each dNTP, 0.24 μM of each primer and 0.5 U of BioTaq DNA polymerase (Bioline, Luckenwalde, Germany). The PCR mixture without cDNA template served as negative control to exclude false positive data.

Cloning strategies. Novel PCR products were cleaned with Wizard SV Gel and PCR Clean-Up Kit (Promega, Mannheim, Germany) and subsequently cloned into pGEM-Teasy vector (Promega) and transformed into *E. coli* K12 JM109 strain. Positive colonies were checked by PCR and agarose-gel electrophoresis. Appropriate colonies were amplified by overnight incubation. Thereafter the plasmids were isolated by the Nucleospin Plasmid Kit (BD Biosciences, Heidelberg, Germany) according to the manufacturer's instructions and sequenced by using a Big-Dye terminator kit on an ABI377 automated sequencer (Perkin-Elmer Instruments GmbH, Rodgau-Jügesheim, Germany, IZKF Core Unit of the University of Leipzig).

For a quantitative *CYP1A1* PCR on Rotorgene (ltf-Labortechnik, Wasserburg/B., Germany) the primer pair h-q*CYP1A1*_for/rev (5'-TCC TGG AGC CTC ATG TAT TTG GTG-3', 5'-TTG TTG TGC TGT GGG GGA TGG TGA-3', respectively) was used amplifying a 196-bp long fragment of exons 4 to 5/6. The PCR was performed at 94˚C for 15 min, then 40 cycles at 94˚C for 30 sec, at 58˚C for 30 sec, at 72˚C for 30 sec. The PCR mixture contained 1.25 mM $MgCl_2$, 200 μM each dNTP, 0.24 μM of each primer, 1 U of Polymerase (HOTFIREPol, Solis BioDyne, Tartu, Estonia) and a 5X concentrate of fluorogenous substrate SYBR green (MoBiTec, Göttingen, Germany). A specific PCR product was present when the melting curve showed a peak at 87.3˚C (negative control at 79.2˚C).

In silico analyses. Computational analyses were carried out using the following databases: UCSC Genome Browser (http://genome.ucsc.edu/), Ensembl Genome Browser (http://www.ensembl.org/), BLAST, dbSNP provided by the National Center for Biotechnology Information (NCBI http://www.ncbi.nlm.nih.gov), ESEfinder (http://rulai.cshl.edu/tools/ESE/) SNP500Cancer (http://snp500cancer.nci.nih.gov/).

Results

To assess human health risk toward environmental pollutants we were interested in frequently occuring polymorphisms

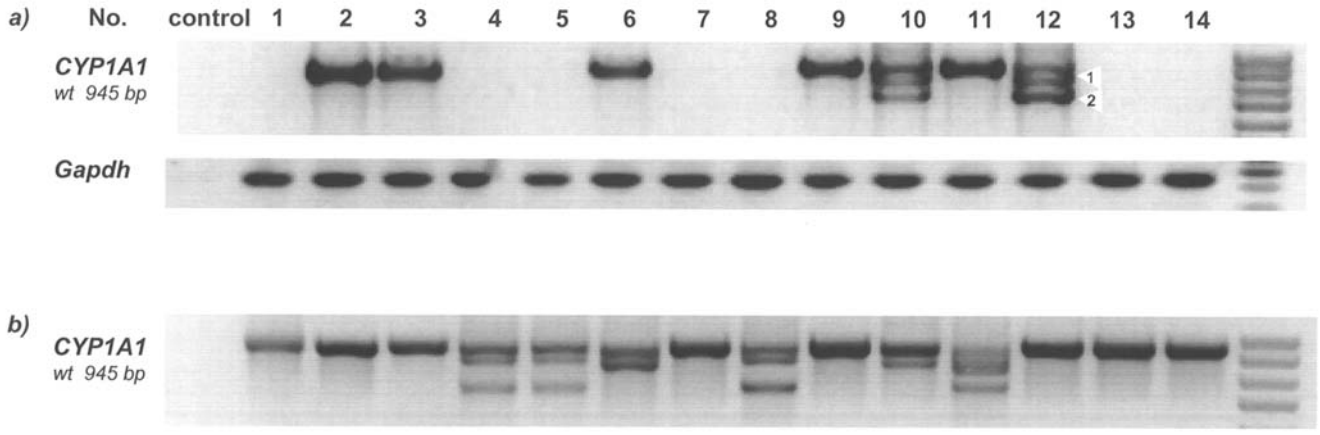

Figure 1. Detection of basal *CYP1A1* long fragment transcripts in leucocytes of representative healthy volunteers using nested reverse-transcription polymerase chain reaction (RT-PCR, primer pairs: outer CYP1A1-for2/-rev1, inner -for3/-rev2). (a) *CYP1A1* mRNA was not detectable in all volunteers. *Gapdh* served as internal positive control. Multiple alternative splicing pattern are present (numbered arrow 1-2 are indicated). (b) Multiple alternative splicing variants are frequently present among volunteers tested positive. One or two additional transcripts are detectable. M, 100 bp DNA ladder, reversed image.

and transcript variations of xenobiotic metabolizing enzymes within epidemiological studies. When focusing on *CYP1A1* mRNA expression, different RT-PCR results were surprisingly obtained by using several *CYP1A1* specific primer pairs. Thus, we performed long fragment RT-PCR approaches on leucocytes of healthy volunteers. A nested long fragment RT-PCR approach on leucocytes provided the expected *CYP1A1* transcript of 945 bp in about 70% (19 out of 27) of the healthy volunteers (Fig. 1a). Approximately 42% (8 out of 19) of the CYP1A1 positive volunteers expressed one or two additional RT-PCR transcripts (Fig. 1b, numbered arrow 1 and 2). Out of the *CYP1A1* positive transcripts about 25% and 75% of volunteers expressed one or two additional transcripts, respectively. The presence or absence of *CYP1A1* transcript and their multiple variants had equal gender distribution. Cloning and sequencing of the two additional transcripts revealed a 84-bp deletion in exon 2 for the 861-bp variant and a 193-bp deletion in exon 2 for the 752-bp variant. The transcript with the 84-bp deletion is identical to the enzymatic active splicing variant of ovary cancer cell lines published under GenBank accession no. AY310359. This splicing variant was frequently expressed in 30% (8 out of 27) of the healthy volunteers who tested positive. The novel 752 bp variant is published under accession number AM233519 by the authors.

To estimate the inducibility of *CYP1A1* in formerly RT-PCR transcript-negative volunteers we performed an *in vitro* stimulation of Cyp1A1 on density separated mononuclear cell (MNC) and granulocytes. Healthy volunteers with an absent *CYP1A1* RT-PCR transcript becomes *CYP1A1* transcript positive for the MNC fraction after *in vitro* stimulation with the CYP1A1-inducer benzo(a)pyrene (B(a)P). *CYP1A1* remained absent in the granulocyte fraction (data not shown). Interestingly, novel splicing variants of *CYP1A1* were induced by stimulation of individual healthy MNC (Fig. 2, numbered arrows 3 and 4). Cloning and sequencing of the new splicing transcripts revealed a 589 bp (365 bp RT-PCR band, Ensembl accession no. AM233517) and 782 bp (163 bp band, Ensembl accession no. AM236047) deletion of the 5-primed part of the exon 2, respectively. In summary, four additional splicing

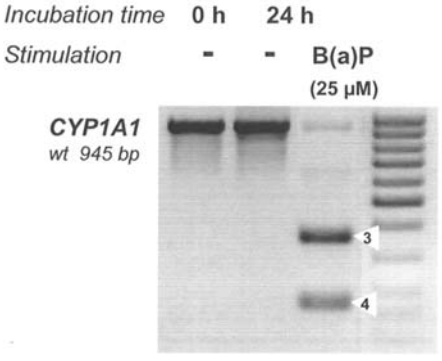

Figure 2. *CYP1A1* mRNA expression of benzo(a)pyrene (B(a)P)-stimulated mononuclear cells (MNC) *in vitro* using nested reverse-transcription polymerase chain reaction (RT-PCR). The B(a)P stimulation additionally induced novel splice variants (numbered arrows 3-4 are indicated) in MNC. M, 100 bp DNA ladder, reversed image.

variants of *CYP1A1* (one known and three novel) become apparent in blood leucocytes of healthy volunteers.

The transcriptional expression of alternative splicing variants of *CYP1A1* in healthy volunteers was compared to splicing pattern in human cancer cell lines established from several tissues. As depicted in Fig. 3, multiple splicing variants were expressed in individual human cancer cell lines (numbered arrows 5-9). Three variants were similar to leucocytes (arrows 6, 7, 9) and two variants represented novel (arrows 5, 8, Ensembl accession no. AM233520 and AM 233518) variants. Human ovary cancer cell lines were found to express an additional transcript besides the firstly described *CYP1A1v* by Leung *et al* (19) (Fig. 3b) in OVCA cells, similar to leucocytes. However, in all cases the expected full length *CYP1A1* transcript was overexpressed compared to the additional splicing variants. Finally, the *CYP1A1* splicing pattern did not differentiate between non-SCLC and SCLC and was not specific for discriminating hepatoma-derived cell lines from immortalized hepatocyte cell lines.

The constitution of the above mentioned additional splicing transcripts of human *CYP1A1* is summarised in Fig. 4.

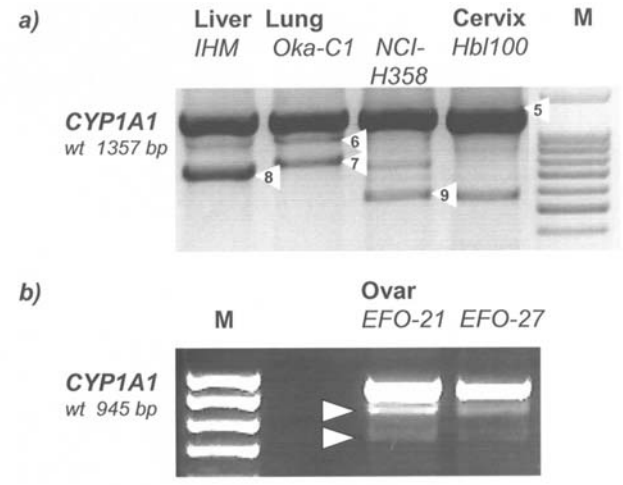

Figure 3. Detection of additional *CYP1A1* splicing variants among human cancer cell lines of different tissues using reverse-transcription polymerase chain reaction (RT-PCR). (a) At least five additional *CYP1A1* transcripts (numbered arrow 5-9 are indicated) are visible after EtBr-stained agarose gel electrophoresis in different human cancer cell lines (reversed image). (b) Both ovary derived cancer cell lines show a multiple pattern similar to leucocytes. M, 100 bp DNA ladder.

The main disturbed splicing events were realised within exon 2. At least four variants with single deletion of 84, 193, 589 and 782 bp in exon 2 were frequently found in blood leucocytes of healthy volunteers (white numbered arrows). These variants were commonly expressed in human cancer cell lines additionally to splicing variants with partial or full retention of intron sequences (grey numbered arrows).

Computational analysis of the splicing variants of *CYP1A1* revealed that deletions within exon 2 cause truncated proteins covering either the NH_2- or the COOH-terminal region of the wild-type protein with one exception. The AY310359 variant having 28 fewer amino acid residues can be translated into an enzymatic active protein. The presence of other protein variants was not examined by the present study. However, theoretical truncated proteins of the COOH-terminal region of the wild-type protein would contain structural qualities necessary for the enzyme such as the cytochrome P450 cysteine heme-iron ligand signature, oxygen binding sites and the substrate binding site at position Val382. Additionally, the COOH-terminal region contains a putative bipartite nuclear localization signal (NLS) which may evoke a nuclear distribution other the endoplasmic reticulum localization of the wild-type enzyme.

The alternative splice sites within exon 2 of human *CYP1A1* gene and their resulting splicing variants were depicted in detail in Fig. 5. Two conservative (GT-AG rule) and two non-canonical splice sites were found. All the four splice sites were located upstream of the 3-primed end of exon 2 in a maximal distance of 192 bp to each other. Alternative splicing variants were performed exclusively between conservative or non-canonical splice sites. Following the GT-AG rule all three possible splicing combinations were transcribed beginning

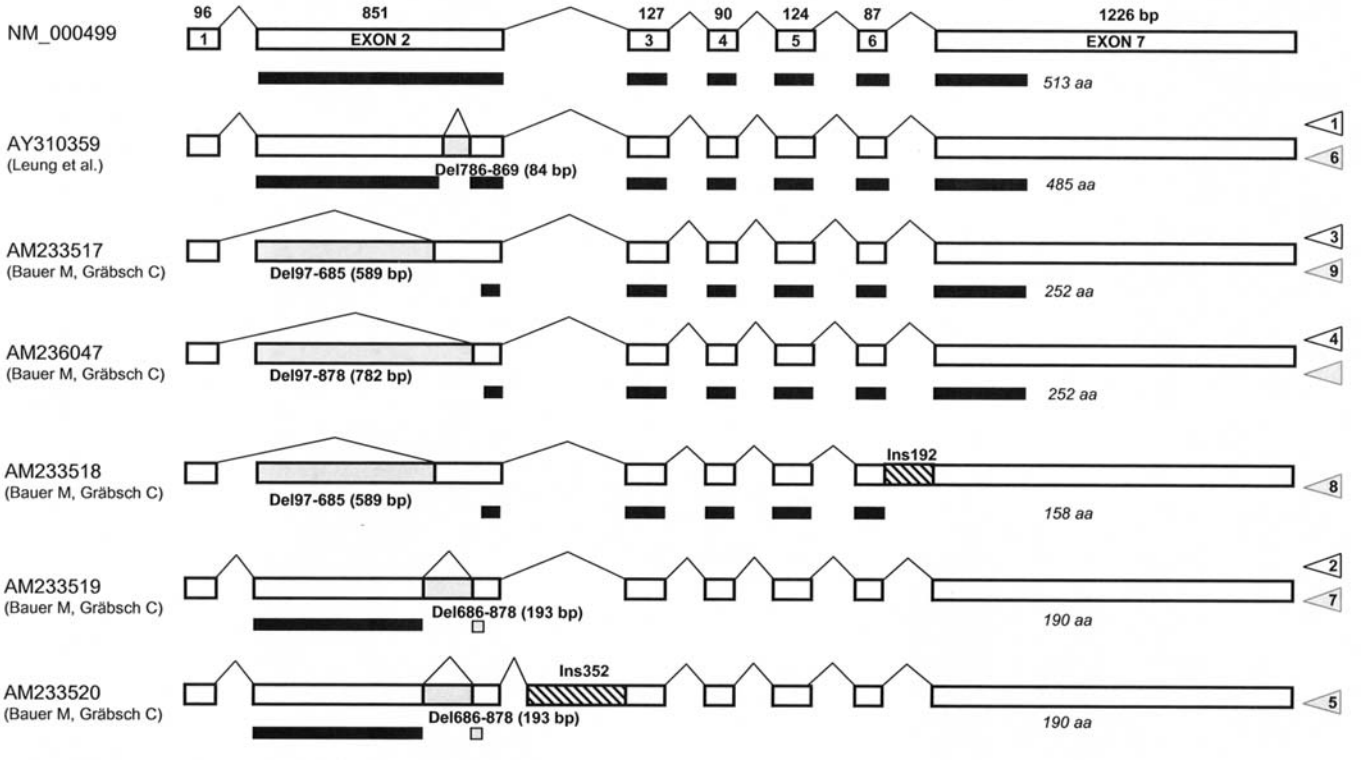

Figure 4. Summarized alternatively spliced transcripts of human *CYP1A1* expressed in human cancer cell lines or leucocytes of healthy volunteers (grey and white numbered arrows, respectively). The splicing variants represent the five novel transcripts described by the authors, and one recently described by Leung *et al* (19). Exon numbering and accession number are given according to EMBL database. The alternative spliced mRNA transcripts and putative in-frame translation products (white and black boxes, respectively) are indicated. Deletions and intron insertions (full insertion, AM233518; partial insertion, AM233520) are depicted (grey and striped boxes, respectively).

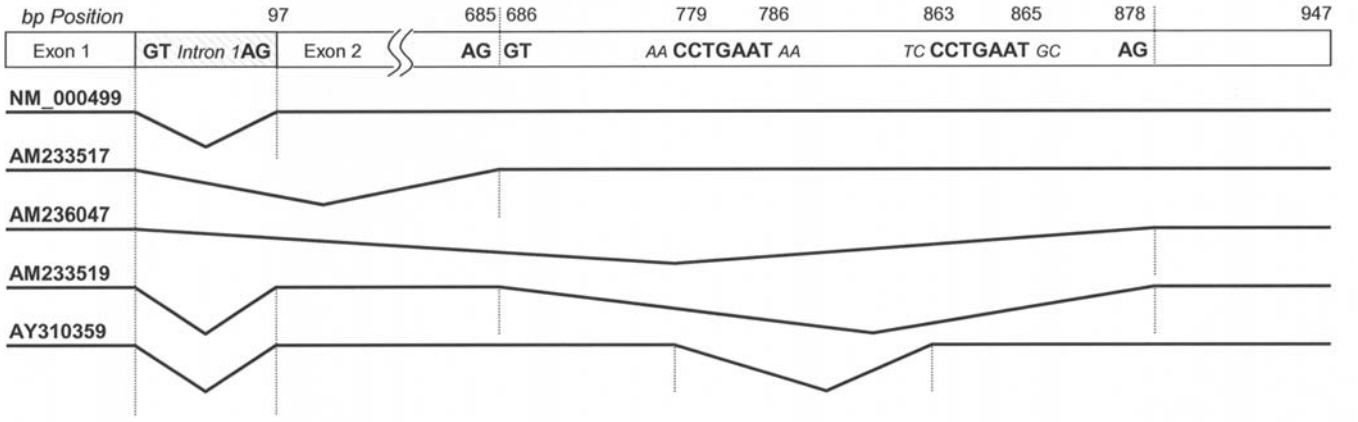

Figure 5. Alternative splice sites within exon 2 of human *CYP1A1* gene and resulting splicing variants. Two conservative (GT-AG rule) and two non-canonical splice sites were found. Following the GT-AG rule all three splicing combinations are transcribed starting from intron 1. In contrast, the non-canonical splice sites form only one splicing transcript.

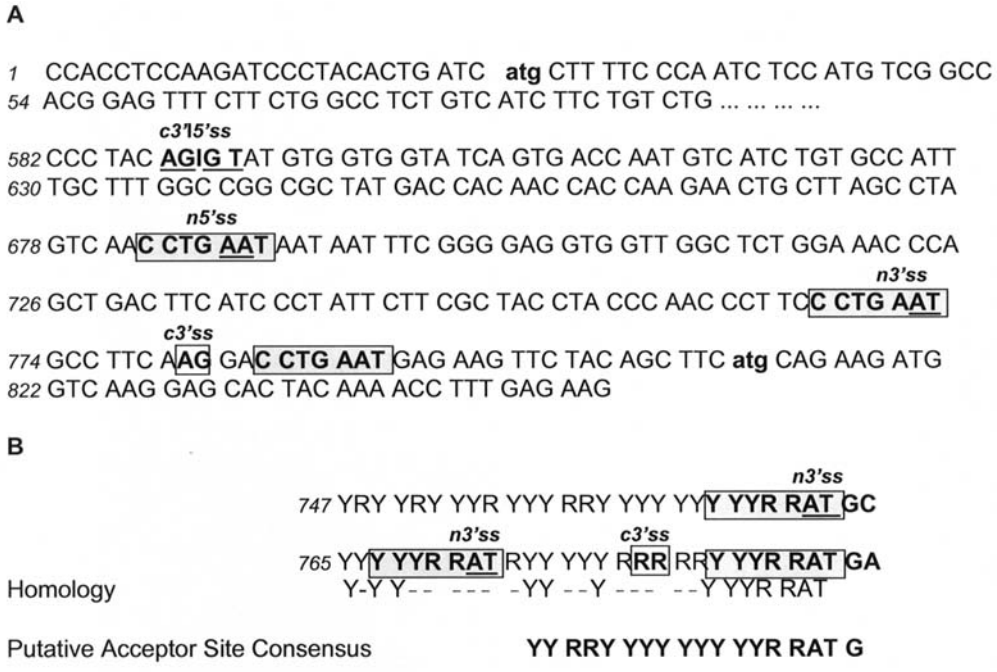

Figure 6. Comparison of sequence homology of two putative non-canonical acceptor splice sites within exon 2 of human *CYP1A1* pre-mRNA (NM-000499). (A) Localization of nested alternative canonical (c3'/5'ss, c3'ss, white boxes) and non-canonical (n5'ss, n3'ss) splice sites is indicated. The grey box highlights a frequent sequence ranging over both confirmed non-canonical splice sites. (B) Upstream sequences of the two putative non-canonical acceptor splice sites are compared by homology in a purin/pyrimidine transformed nucleotide code. A proposed acceptor site consensus is indicated due to an absent homology of the upstream sequence of the two repeats. The polypyrimidine tract (Py) at the 5' end of the n3'ss seems to be necessary for the non-canonical splicing at this splicing site and should be contained in an acceptor site consensus for the non-canonical AA-AT splice pair.

from intron 1. In contrast, only one splicing transcript was released employing the two non-canonical splice sites.

Discussion

Cytochrome P450 1A1 (CYP1A1) plays an important role in the defense against environmental pollutants. The chronic overexpression of CYP1A1 independently of gender represents a risk for human cancer and has been investigated in several studies (32). Recently, in ovarian cancer cell lines a novel aberrant CYP1A1 protein was described to have an additional impact in tumour initiation and progression (19). Surprisingly, focusing on *CYP1A1* transcript pattern within an epidemiological study and in toxicological *in vitro* approaches we were able to extend the physiological appearance of multiple transcripts of human CYP1A1 by the present study.

Leung *et al* (19) have demonstrated that the *CYP1A1wt* and the variant *CYP1A1v* mRNA were concomitantly expressed both at minimal level in normal human ovarian surface epithelial cells (HOSE) and at overexpressed level in human ovarian cancer cell lines (OVCA). When HOSE cell lines were

stably transfected with the wild-type *CYP1A1* transcript then the alternative splice variant, *CYP1A1v* becomes concomitantly overexpressed in addition to the expected wild-type transcript. This finding suggests an exclusive role of a tissue-specific spliceosome machinery in generating multiple *CYP1A1* transcripts. *CYP1A1*-specific inducer, such as benz(a)pyrene, failed to interrupt the multiple transcript pattern.

Interestingly, the *CYP1A1v* mRNA released an enzymatic active CYP1A1v protein which was exclusively localized in cell nucleus and mitochondria. The unusual localization of CYP1A1v implicates its role in ovarian cancer initiation and progression by direct release of genotoxic active metabolites within the nucleus. However, an expected higher risk for neoplastic transformation of HOSE by a single overexpressed nuclear localized CYP1A1v, compared to single overexpressed endoplasmatic localized CYP1A1wt, has not been supported *in vitro*. Unfortunately, co-transfection experiments with *CYP1A1wt* and *CYP1A1v*, which would have been represented the *in vivo* concomitant expression of multiple *CYP1A1* transcripts, were not performed. In addition, considering that the spliceosome machinery prefers to release *CYP1A1wt* mRNA, as indicated by semi-quantitative RT-PCR, and that only a minority of nuclei could have been immunohistochemically stained on human ovarian tumours, the CYP1A1wt protein should possess the dominant role over CYP1A1v in CYP1A1 overexpressed human tumours.

The human CYP1A1 gene consists of 7 exons according to the NCBI MapViewer and UCSC Genome database. In the Ensembl Genome database the released transcript sequence for human CYP1A1 (Ensembl release 40- Aug 2006) does not conform with the NCBI/UCSC entries. Sequence data of the first exon are missing and therefore the first numbered exon of the Ensembl database represent exon 2 of the NCBI/UCSC databases. Considering the full length transcript information of the present study we retained the seven exon numbering of the *CYP1A1* transcript.

In the present study five novel *CYP1A1* transcripts are described, additionally to the recently described *CYP1A1v*. Similar to alternative *CYP2E1* and *GSTT1* pre-mRNA splicing (33,34), the exon 2 is the most affected exon regarding alternative transcripts of human *CYP1A1*. It additional contains 2 nested splice site pairs. One outer canonical and one inner non-canonical splice site pair. Within this exon the core 5'-canonical splice site (5'ss, AGGU) serves either as a donor splice site (5'ss) by use of the AG or as an acceptor splice site (3'ss) by use of the GU. Because of the most conserved AT dinucleotides at the -2 and -1 exonic position of a dual utilized canonical 5'ss, the observed splicing variants could have been predicted for such pre-mRNA types (35). Similar splicing events at a dual utilized 5'ss were described for the interferon regulatory factor-3 (IRF-3) (36). However, the alternative splicing of the *IRF-3* pre-mRNA distinguished from that of *CYP1A1* by using exclusively one dual utilized exonic splice site. As a result, the whole up- or downstream to the splice site exonic sequence is spliced, different from a spliced cryptic intron for *CYP1A1* pre-mRNA. In summary, the exon 2 of CYP1A1 can be considered as assembled from two exons and one nested retained cryptic intron.

The use of alternative and inefficient splice sites may be influenced by competitive binding of serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNP) as some of the factors which mediate splicing regulatory activities. The mechanism by which splice sites are used appears to be dynamically regulated, at least in part, by the relative ratio of hnRNP A1 to SR proteins in the nucleus (37). Among the physiologically occured multiple transcript pattern of *CYP1A1* the wild-type transcript is permanently present and overexpressed over the concomitant alternative transcripts. Therefore, the dynamic balance between splicing enhancer and inhibitors seems to be shifted to the site of preferentially expressed wild-type transcript.

In contrast, compared to additional multiple transcripts the wild-type transcript has become underrepresented after exposure of high dose of B(a)P of individual human leucocytes. At least two mechanisms has to be considered to explain this changed splicing pattern. At first, metabolism of B(a)P results in formation of the ultimate carcinogenic form of benzo(a)pyrene, the 7,8 diol 9,10 epoxide. Covalently bound to DNA, this electrophilic metabolite should be able to block target sequences for binding of splicing factors. Secondly, genotoxic stress may induce phosphorylation of definite splicing factors and results in the translocation of the splicing factor from the nucleus into the cytoplasm (37). Thus, an altered ratio between splicing enhancer and inhibitors may also cause an altered splice site selection as evoked for *CYP1A1* after a genotoxic impact by B(a)P exposure.

An aberrant splicing of either alternatively or constitutively spliced genes can also be triggered by mutations in the spliced gene by influencing binding affinity of splicing proteins to the DNA (38). As a profound consequence human inherited disorders can be manifested (39,40). Regarding the frequent appearence of multiple *CYP1A1* splicing transcripts in leucocytes of healthy volunteers (42%), in the first instance, it appears unlikely that their presence has been provoked by mutations. However, to prove the possibility of underlying mutations we performed a single nucleotide polymorphism (SNP) database analysis for *CYP1A1* as well as SNP detection for the most prominent Ile/Val (refSNP ID rs1048943) and MspI (refSNP ID rs4646903) variations. At present, 37 SNPs are registered for the genomic sequence of human *CYP1A1* gene. Allele-specific frequency data are disposable from 19 SNPs. Of these, one SNP (refSNP ID rs2606345) in intron 1 coincides with the expected allele-specific frequencies for a Caucasian population. Due to a 70% heterozygosity the frequencies of the two possible alleles of this SNP are 45 and 55%. Thus a genomic variation in form of a SNP has not to be excluded as a mechanism of splicing variation for the human *CYP1A1* gene. Ile/Val and MspI variations failed to correlate with the multiple splicing pattern (data not shown).

A sequence of 7 nucleotides (*CCTGAAT*) ranges over the non-canonical both 5'ss (n5'ss) and 3'ss (n3'ss, Fig. 6A). None of the possible dinucleotide splice pair combinations follow either a canonical GT-AG (having 99.24% of annotated mammalian genes) (41) or the most common non-canonical GC-AG (0.69%) and AT-AC (0.05%) rule. It therefore remains speculative which of the non-canonical dinucleotide combinations finally serve as splice junction.

However, following the dataset given by Burset *et al* (41) and the suggestion of Leung *et al* (19) the AA-AT splice pair should be the most likely non-canonical splice pair.

A third in-frame *CCTGAAT* repeat is located 11 nucleotides downstream to the non-canonical 3'ss within the exon 2 of *CYP1A1*. However, the AT dinucleotide of this repeat seems not to be recognized by the splicing machinery as a 3'ss. Its near location to the non-canonical 3'ss may give an indirect evidence about splicing acceptor consensus for the 3'ss of the AA-AT splicing pair.

A spliceosome assembly depends on well-defined conserved consensus sites upstream of a 5'ss and downstream of a 3'ss. Definite splice site consensus for the rare AA-AT splice pair were not postulated to date. In general, the consensus for the 3'ss spans three elements: the branchpoint sequence, the polypyrimidine tract (Py), and the four-nucleotide sequence at the 3' splice site. Even though these sequence elements conform to a consensus, there is significant variability (42). Comparing the upstream sequences of the *CCTGAAT* repeat elements at the two possible acceptor splice sites, it reveals that the most downstream repeat lacks a Py. It suggested, that a 9 Py of the n3'ss may be necessary for the spliceosome assembly of rare AA-AT splice pairs. As a first approximation, the proposed by the present study non-canonical acceptor site consensus: RYYYYYYYYYRRATIG (Fig. 6B) has been supported by the similar acceptor site consensus for the most prominent canonical GT-AG splice pair: YYYYYYYYYNCAGIG (43).

Furthermore, blocking of alternative splice sites for spliceosome proteins by exon splicing enhancer proteins (ESE) may constitute another conceivable mechanism to omit the third repeat as a non-canonical 3'ss. However, stringent evidence could not be proposed after analyzing computational data of putative DNA binding sites for splicing factors by the ESEfinder software (44).

In conclusion, we here described common multiple human *CYP1A1* transcripts in healthy volunteers as well as multiple transcripts exclusively found in human tumour cell lines. The frequent appearence of the common multiple transcripts has led us to discuss their presence as a physiological feature rather than a risk for cancer. The individual differences in alternative *CYP1A1* transcript pattern will be a focus in further epidemiological studies to clarify whether the pattern represents a signature of a specific human exposure.

References

1. Kim JH, Stansbury KH, Walker NJ, Trush MA, Strickland PT and Sutter TR: Metabolism of benzo[a]pyrene and benzo[a]pyrene-7,8-diol by human cytochromeP450 1B1. Carcinogenesis 19: 1847-1853, 1998.
2. Shimada T, Tsumura F and Yamazaki H: Prediction of human liver microsomal oxidations of 7-ethoxycoumarin and chlorzoxazone with kinetic parameters of recombinant cytochrome P-450 enzymes. Drug Metab Dispos 27: 1274-1280, 1999.
3. Neber DW, Dalton TP, Okey AB and Gonzalez FJ: Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. J Biol Chem 279: 23847-23850, 2004.
4. Roberts-Thomson SJ, McManus ME, Tukey RH, Gonzalez FF and Holder GM: The catalytic activity of four expressed human cytochrome P450s towards benzo[a]pyrene and the isomers of its proximate carcinogen. Biochem Biophys Res Commun 192: 1373-1379, 1993.
5. Nebert DW, Roe AL, Dieter MZ, Solis WA, Yang Y and Dalton TP: Role of the aromatic hydrocarbon receptor and [Ah] gene battery in the oxidative stress response, cell cycle control, and apoptosis. Biochem Pharmacol 59: 65-85, 2000.
6. Iwanari M, Nakajima M, Kizu R, Hayakawa K and Yokoi T: Induction of CYP1A1, CYP1A2, and CYP1B1 mRNAs by nitropolycyclic aromatic hydrocarbons in various human tissue-derived cells: chemical-, cytochrome P450 isoform-, and cell-specific differences. Arch Toxicol 76: 287-298, 2002.
7. Mollerup S, Berge G, Baera R, Skaug V, Hewer A. Phillips DH, Stangeland L and Haugen A: Sex differences in risk of lung cancer: expression of genes in the PAH bioactivation pathway in relation to smoking and bulky DNA adducts. Int J Cancer 119: 741-744, 2006.
8. Kim JH, Sherman ME, Curriero FC, Guengerich FP, Strickland PT and Sutter TR: Expression of cytochromes P450 1A1 and 1B1 in human lung from smokers, non-smokers, and ex-smokers. Toxicol Appl Pharmacol 199: 210-219, 2004.
9. Zhang KL, Ma JX, Chen XY, Sun Y, Kong QY, Liu J and Li H: Frequent CYP1A1 expression in gastric cancers and their related lesions. Oncol Rep 12: 1335-1340, 2004.
10. Nagara NS, Beckers S, Mensah JK, Waigel S, Vigneswaran N and Zacharias W: Cigarette smoke condensate induces cytochromes P450 and aldo-keto reductases in oral cancer cells. Toxicol Lett 165: 182-194, 2006.
11. Parkinson A: Biotransformation of xenobiotics. In: Casarett and Doull's Toxicology, The Basic Science of Poisons. 6th edition. Klaassen CD (ed). McGraw-Hill Professional, 2001.
12. Goda R, Nagai D, Akiyama Y, Nishikawa K, Ikemoto I, Aizawa Y, Nagata K and Yamazoe Y: Detection of a new N-oxidized metabolite of flutamide, N-[4-nitro-3-(trifluoromethyl)phenyl]hydroxylamine, in human liver microsomes and urine of prostate cancer patients. Drug Metab Dispos 34: 6828-6835, 2006.
13. Haas S, Pierl C, Harth V, Pesch B, Rabstein S, Bruning T, Ko Y, Hamann U, Justenhoven C, Brauch H and Fischer HP: Expression of xenobiotic and steroid hormone metabolizing enzymes in human breast carcinomas. Int J Cancer 23: [Epub ahead of print] 2006.
14. Iscan M, Klaavuniemi T, Coban T, Kapucuoglu N, Pelkonen O and Raunio H: The expression of cytochrome P450 enzymes in human breast tumours and normal breast tissue. Breast Cancer Res Treat 70: 47-54, 2001.
15. Hellmold H, Rylander T, Magnusson M, Reihner E, Warner M and Gustafsson JA: Characterization of cytochrome P450 enzymes in human breast tissue from reduction mammaplasties. J Clin Endocrinol Metab 83: 886-895, 1998.
16. Parkinson A and Hurwitz A: Omeprazole and the induction of human cytochrome P-450: a response to concerns about potential adverse effects. Gastroenterology 100: 1157-1164, 1991.
17. Furukawa M, Nishimura M, Ogino D, Chiba R, Ikai I, Ueda N, Naito S, Kuribayashi S, Moustafa MA, Uchida T, Sawada H, Kamataki T, Funae Y and Fukumoto M: Cytochrome p450 gene expression levels in peripheral blood mononuclear cells in comparison with the liver. Cancer Sci 95: 520-529, 2004.
18. Ali S, El-Rayes BF, Heilbrun LK, Sarkar FH, Ensley JF, Kucuk O and Philip PA: Cytochrome p450 and glutathione transferase expression in squamous cell cancer. Clin Cancer Res 10: 4412-4416, 2004.
19. Leung YK, Lau KM, Mobley J, Jiang Z and Ho SM: Over-expression of cytochrome P450 1A1 and its novel spliced variant in ovarian cancer cells: alternative subcellular enzyme compartmentation may contribute to carcinogenesis. Cancer Res 65: 3726-3734, 2005.
20. Little CD, Nau MM, Carney DN, Gazdar AF and Minna JD: Amplification and expression of the c-myc oncogene in human lung cancer cell lines. Nature 306: 194-196, 1983.
21. Turner AM, Zsebo KM, Martin F, Jacobsen FW, Bennett LG and Broudy VC: Nonhematopoietic tumor cell lines express stem cell factor and display c-kit receptors. Blood 80: 374-381, 1992.
22. Lieber M, Smith B, Szakal A, Nelson-Rees W and Todaro G: A continuous tumor-cell line from a human lung carcinoma with properties of type II alveolar epithelial cells. Int J Cancer 17: 62-70, 1976.
23. Bepler G, Koehler A, Kiefer P, Havemann K, Beisenherz K, Jaques G, Gropp C and Haeder M: Characterization of the state of differentiation of six newly established human non-small-cell lung cancer cell lines. Differentiation 37: 158-171, 1988.

24. Asahi Y, Kubonishi I, Imamura J, Kamioka M, Matsushita H, Furihata M, Ohtsuki Y and Miyoshi I: Establishment of a clonal cell line producing granulocyte colony-stimulating factor and parathyroid hormone-related protein from a lung cancer patient with leukocytosis and hypercalcemia. Jpn J Cancer Res 87: 451-458, 1996.
25. Heidtmann HH, Salge U, Havemann K, Kirschke H and Wiederanders B: Secretion of a latent, acid activatable cathepsin L precursor by human non-small cell lung cancer cell lines. Oncol Res 5: 441-451, 1993.
26. Takaki T: An epithelial cell line (KNS-62) derived from a brain metastasis of bronchial squamous cell carcinoma. J Cancer Res Clin Oncol 96: 27-33, 1980.
27. Gazdar AF, Linnoila RI, Kurita Y, Oie HK, Mulshine JL, Clark JC and Whitsett JA: Peripheral airway cell differentiation in human lung cancer cell lines. Cancer Res 50: 5481-5487, 1990.
28. Brower M, Carney DN, Oie HK, Gazdar AF and Minna JD: Growth of cell lines and clinical specimens of human non-small cell lung cancer in a serum-free defined medium. Cancer Res 46: 798-806, 1986.
29. Ellison M, Woodhouse D, Hillyard C, Dowsett M, Coombes RC, Gilby ED, Greenberg PB and Neville AM: Immunoreactive calcitonin production by human lung carcinoma cells in culture. Br J Cancer 32: 373-379, 1975.
30. Aden DP, Fogel A, Plotkin S, Damjanov I and Knowles BB: Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line. Nature 282: 615-616, 1979.
31. Simon WE, Albrecht M, Hansel M, Dietel M and Holzel F: Cell lines derived from human ovarian carcinomas: growth stimulation by gonadotropic and steroid hormones. J Natl Cancer Inst 70: 839-845, 1983.
32. Denissenko MF, Pao A, Tang M and Pfeifer GP: Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. Science 274: 430-432, 1996.
33. Bauer M, Herbarth O, Aust G and Graebsch C: Molecular cloning and expression of novel alternatively spliced cytochrome P450 2E1 mRNAs in humans. Mol Cell Biochem 280: 201-207, 2005.
34. Bauer M, Herbarth O, Aust G, Hengstler JG, Graebsch C and Schmuecking E: Expression pattern and novel splicing variants of Glutathione S-transferase isoenzymes of human lung and hepatocyte cell lines. Cell Tissue Res 324: 423-432, 2006.
35. Hastings ML and Krainer AR: Pre-mRNA splicing in the new millennium. Curr Opin Cell Biol 13: 302-309, 2001.
36. Karpova AY, Howley PM and Ronco LV: Dual utilization of an acceptor/donor splice site governs the alternative splicing of the IRF-3 gene. Genes Dev 14: 2813-2818, 2000.
37. Van der Houven van Oordt W, Diaz-Meco MT, Lozano J, Krainer AR, Moscat J and Caceres JF: The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNPA1 and modulates alternative splicing regulation. J Cell Biol 149: 307-316, 2000.
38. Philips AV and Cooper TA: RNA processing and human disease. Cell Mol Life Sci 57: 235-249, 2000.
39. Ars E, Serra E, Garcia J, Kruyer H, Gaona A, Lazaro C and Estivill X: Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. Hum Mol Genet 9: 237-247, 2000.
40. Teraoka SN, Telatar M, Becker-Catania S, Liang T, Onengut S, Tolun A, Chessa L, Sanal O, Bernatowska E, Gatti RA and Concannon P: Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. Am J Hum Genet 64: 1617-1631, 1999.
41. Burset M, Seledtsov IA and Solovyev VV: Analysis of canonical and non-canonical splice sites in mammalian genomes. Nucleic Acids Res 28: 4364-4375, 2000.
42. Burge CB, Tuschl T and Sharp PA: Splicing of precursors to mRNAs by the spliceosomes. In: The RNA world. Gesteland RF, Cech TR and Atkins JF (eds). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp525-560, 1999.
43. Burset M, Seledtsov IA and Solovyev VV: SpliceDB: database of canonical and non-canonical mammalian splice sites. Nucleic Acids Res 29: 255-259, 2001.
44. Cartegni L, Wang J, Zhu Z, Zhang MQ and Krainer AR: ESEfinder: a web resource to identify exonic splicing enhancers. Nucleic Acids Res 31: 3568-3571, 2003.