



US 20040121327A1

(19) **United States**
 (12) **Patent Application Publication** (10) **Pub. No.: US 2004/0121327 A1**
Manns et al. (43) **Pub. Date: Jun. 24, 2004**

(54) **METHOD FOR THE PREDICTION OF THE RISK POTENTIAL FOR CANCEROUS DISEASES AND INFLAMMATORY INTESTINAL DISEASES AND CORRESPONDING TESTS**

Publication Classification

(51) **Int. Cl.⁷** **C12Q 1/68; C07H 21/04**
 (52) **U.S. Cl.** **435/6; 536/25.4**

(76) Inventors: **Michael Manns**, Isernhagen (DE);
Christian Strassburg, Hannover (DE)

(57) **ABSTRACT**

Correspondence Address:
WHITHAM, CURTIS & CHRISTOFFERSON, P.C.
11491 SUNSET HILLS ROAD
SUITE 340
RESTON, VA 20190 (US)

The invention relates to a method for the prediction of the risk potential and/or diagnosis of cancerous diseases or inflammatory intestinal diseases, whereby a DNA sample is tested for the presence of polymorphic UGT1A7 allele. A positive result for a mutation is a positive indication of a sensitivity to cancerous diseases. A prediction of sensitivity to an inflammatory intestinal disease can similarly be made. A PCR amplification of the exon 1, by means of the DNA sample with subsequent sequence analysis is carried out in the method and the determined sequence compared with that of the wild type and the polymorphic allele. The presence or lack of mutations is monitored by means of sequencing the corresponding cDNA using automated fluorescent dye sequencing. The test arrangement for said method requires genetic detection reagents, namely the required primer or cDNAs, on a stationary support in a pre-prepared arrangement or sequence for reading off the results. The recombinant UGT1A7 enzymes are also used for therapeutic purposes.

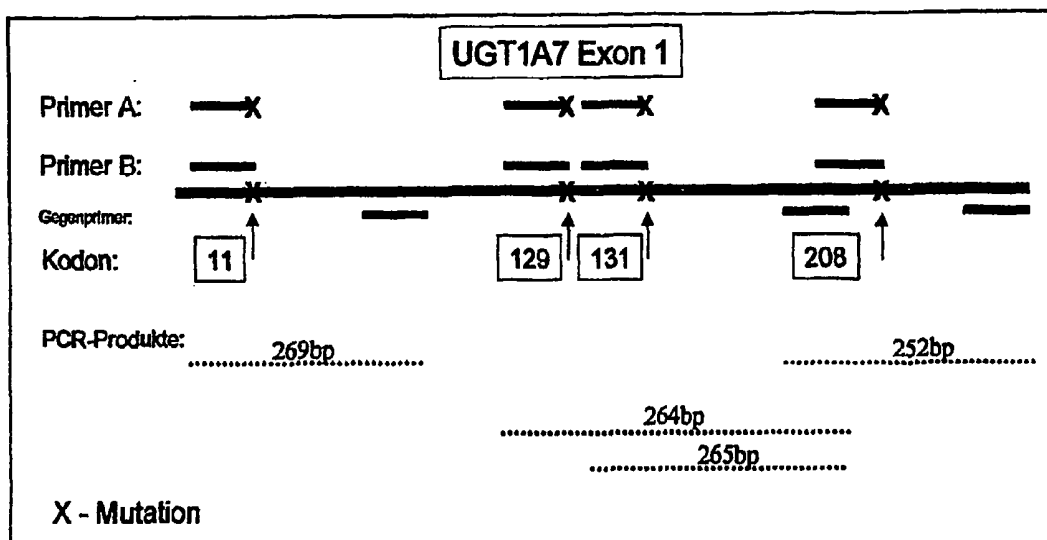
(21) Appl. No.: **10/250,508**

(22) PCT Filed: **Jan. 3, 2002**

(86) PCT No.: **PCT/DE02/00003**

(30) **Foreign Application Priority Data**

Jan. 5, 2001 (DE)..... 101 00 238.6



Figur 1

Allele		NC	CRC	IBD
UGT1A7*1	wildtype	48%	4%	6%
UGT1A7*2	+/+	0%	0%	0%
UGT1A7*2	+/-	5%	19%*	3%
UGT1A7*3	+/+	16%	0%	10%
UGT1A7*3	+/-	18%	23%	10%
UGT1A7*4	+/+	7%	23%*	26%*
UGT1A7*4	+/-	3%	19%*	16%*
UGT1A7*3/ UGT1A7*4	compound heterozygous	3%	12%	16%*
UGT1A7*4 (+/+) and UGT1A7*3/UGT1A7*4		10%	35%*	50%*

Figur 2

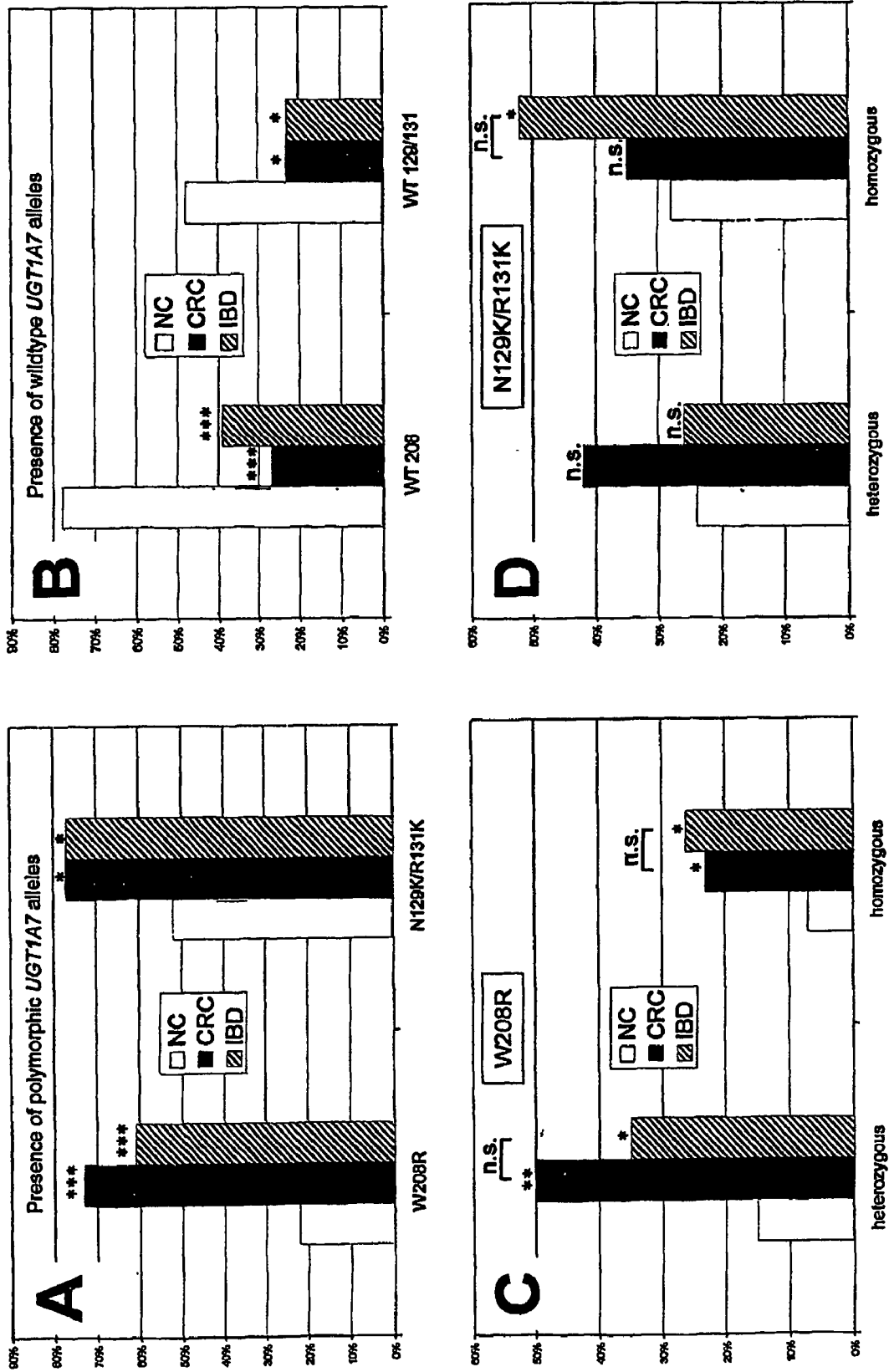
allele	codon no.	10 11 12	128 129 130 131 132	207 208 209	description
UGT1A7*1		CTT CCC CTA L P L	TTT AAT GAC CGA AAA F N D R K	GTA TGG AAC V W N	WT
UGT1A7*2		CTT CCA CTA L P L	TTT AAT GAC CGA AAA F N D R K	GTA CGG AAC V R N	W208R
UGT1A7*3		CTT CCC CTA L P L	TTT AAG GAC AAA AAA F K D K K	GTA TGG AAC V W N	N129K/R131K
UGT1A7*4		CTT CCA CTA L P L	TTT AAG GAC AAA AAA F K D K K	GTA CGG AAC V R N	N129K/R131K/W208R

Figur 3

aminoacid no.	aa 7 - 15	aa 127 - 134	aa 205 - 213
UGT1A1	GGRP-LVLG---	LLHNKELM---	QRVKNMLIA-
UGT1A3	VPLPWLATG---	LLHNEALI---	QRVKNMLYP-
UGT1A4	VPLPQLATG---	LLHNEALI---	QRVKNMLYP-
UGT1A5	VPLPRLATG---	LLHNEALI---	QRVKNMLYP-
UGT1A6	RSEQRISAG---	LLQDRDTL---	QRVANFLVN-
UGT1A7	TGILLPLYVC---	LFNDRKLV---	ERVWNHIMH-
UGT1A8	TSPIPLCVS---	LFNDRKLV---	ERVNRHIMH-
UGT1A9	TSPLPLCVC---	LFKDKKLV---	ERVNRHIMH-
UGT1A10	DQPRSEMCV---	LFNDRKLV---	ERVWNHIVH-

UGT1A7 polymorphisms:	N129K/R131K	W208R
-----------------------	-------------	-------

Figur 4



$p < 0,0001$ ***, $p < 0,001$ ** , $p < 0,05$ *

Figur 5

METHOD FOR THE PREDICTION OF THE RISK POTENTIAL FOR CANCEROUS DISEASES AND INFLAMMATORY INTESTINAL DISEASES AND CORRESPONDING TESTS

[0001] The invention relates to a method for predicting the potential risk of carcinomas and inflammatory bowel diseases and for diagnosing these disorders. The invention primarily relates to a method for estimating the potential risk of colorectal carcinomas. The invention further relates to diagnostic tests on DNA samples from an individual to be investigated, and to the use of new polymorphic forms of the UGT1A7 gene for the metabolic characterization of medicaments, especially of tumor therapeutic agents.

[0002] The current assumption is that the risk of cancer is determined by a genetic predisposition and by environmental effects including exposure to carcinogenic substances.

[0003] It has therefore frequently been assumed that there is an association between genetic polymorphisms of carcinogen-metabolizing enzymes and the development of a cancer. However, specific, statistically confirmed associations are difficult to find because the metabolizing cannot be equated with a detoxification. For predicting the risk of cancer it is therefore insufficient just to detect genetic modifications; on the contrary, it is necessary to know that the change leads to clearly adverse metabolic consequences.

[0004] A further problem in finding such an association is that the human metabolism is very complex and an enormous number of substances, including metabolites, proteins and in particular also enzymes, might be suitable markers for a risk of cancer. This also applies to colonic or colorectal carcinoma (CRC), one of the commonest types of cancer in the western world. The human intestine is a large organ influenced by many types of body processes and also the diet. It was therefore necessarily doubtful whether an unambiguous estimation of the risk on the basis of a genetic predisposition is in fact possible.

[0005] A large number of enzyme classes are involved in intestinal metabolism, including those responsible for metabolizing foreign substances.

[0006] Human uridine diphosphate (UDP) 5'-glucuronosyltransferase (UGT) is a large class of enzymes bringing about the glucuronidation of numerous substrates. Within this super family, three families with particular tasks have been identified, UGT1, UGT2 and UGT8. Numerous isoforms are distinguished among the enzymes of the UGT1 family alone. In the last 10 years, 5 UGT isoforms of the 1A family (UGT1A1, 1A3, 1A4, 1A6 and 1A9) have been discovered in the liver and cloned, including the bilirubin UGT isoform UGT1A1. Three extrahepatic UGT1A gene products have been isolated from the biliary tract (UGT1A10), stomach (UGT1A7) and colon (UGT1A8).

[0007] The proteins encoded on the human UGT1A7 gene locus serve, through the glucuronidation, the detoxification of a large number of endogenous and exogenous (xenobiotic, i.e. compounds of industrial origin which are difficult to degrade) compounds in the human body. They are proteins of the so-called phase 2 metabolism in humans, which includes for example acetylation, sulfation and glucuronidation. UGT1A7, which was described as transcript for the first time in 1996, has been demonstrated to detoxify poly-

cyclic hydrocarbons and heterocyclic amines. These substances are also regarded as carcinogens.

[0008] WO 00/06776 discloses the identification of genetic polymorphisms in human UGT2B4, UGT2B7 and UGT2B15 genes which modify UGT2B activity. UGT2 enzymes are involved inter alia in steroid metabolism. Nucleic acids which contain the polymorphic UGT2 sequences have been used to screen patients for an altered metabolism of UGT2B substrates, possible drug-drug interactions and adverse effects/side effects, and for diseases derived from exposure to environmental or occupational poisons. Animal, cell and in vitro models of drug metabolisms has [sic] been established with the aid of the nucleic acids.

[0009] Genetic polymorphisms in the human UGT1 gene which modify the UGT1-dependent drug metabolism were identified in WO 99/57322. Once again, the polymorphic sequences were used to screen patients for an altered metabolism of UGT1 substrates, possible drug-drug interactions and adverse effects/side effects, and for diseases derived from exposure to environmental or occupational poisons. In turn, animal, cell and in vitro models of drug metabolisms were established with the aid of the nucleic acids.

[0010] In addition, functional consequences of a genetic modification of the human UGT1A7 gene have already been described in "C. Guillemette, J. K. Ritter, D. J. Auyeung, F. K. Kessler, D. E. Housman; "Struktural [sic] hetroogeneity [sic] at the UDP glucuronosyltransferase 1 locus: functional consequences of three novel missense mutations in the human UGT1A7 gene", Pharmacogenetics, 2000, 10: 629-644". It is also suggested therein that UGT1A7 has a possible role in the detoxification and elimination of carcinogenic products in the lung. UGT1A7 is expressed in the lung but not in the human large bowel.

[0011] The object of the invention was to find a specific association between a risk of cancer, especially of colonic carcinoma (colorectal carcinoma CRC) and genetic dispositions and to provide a test for detecting this disposition.

[0012] This has been achieved by finding new polymorphic UGT1A7 alleles which have been named UGT1A7*2, UGT1A7*3 and UGT1A7*4 and for which it is possible to show with statistically astonishing relevance an association with the occurrence of a colorectal carcinoma (CRC) and moreover with particular inflammatory bowel disorders (IBD; Crohn's disease) and with the types of cancer pancreatic, hepatic, gastric and esophageal cancer. This result is surprising inasmuch as UGT1A7 is not expressed in the large bowel, i.e. not in the diseased region. It is astonishing and has not been found in previous investigations that a genetic predisposition does not depend on expression in the relevant organ but is also possible in a spatially separate manner. It is evident from this that the genetic modification must have a global and not just an organ-specific effect. It has been possible to show such an association here for the first time.

[0013] The genetic polymorphisms are those of the germ line and not somatic mutations of the tumor. The underlying analyses were carried out on genomic DNA from lymphocytes of patients with tumors (CRC), inflammatory bowel disease (IBD) and reference subjects not having these dis-

orders. The result of analysis of a total of 111 individuals is that the polymorphisms in the individual codons do not occur independently of one another but can be inherited only in combination.

[0014] UGT1A7*2 is characterized by a silent mutation at codon 11 with a change from C to A at position 3 (CCC to CCA). This mutation is located in the signal peptide domain of the endoplasmic reticulum. This is associated with a second mutation at codon 208, specifically W208R, a change from T to C, which leads to a non-conservative replacement of an aromatic tryptophan by a positively charged arginine residue. The polymorphism at codon 208 is located in a sequence region which is conserved not just in the first exons of UGT1A7-10 but in all members of the UGT1A7 family. The presence of an RV-N sequence motif between amino acids 206 and 209 is common to all UGT1A proteins. The W208R exchange in UGT1A7 represents the wild-type sequence of UGT1A8 and UGT1A9. The applicant's investigations show that the polymorphisms at codon 11 and at codon 208 do not occur singly (in a study on 111 people). Characterization of individuals heterozygous at both positions, and the fact that it was not possible to find any individually homozygous pattern at codon 11 or at codon 208, suggests that both exchanges are located on the UGT1A7*2 allele.

[0015] UGT1A7*3 comprises two mutations. The first is a T to G change at codon 129, resulting in a conservative codon exchange from asparagine to lysine (N129K). The second is evident at codon 131 in a double change from C to A at position 1 and G to A at position 2 of the codon, leading to an arginine to lysine exchange (R131K). This polymorphism likewise affects highly conserved sequences of the UGT protein. The leucine-127 which is present in all UGT1A7 proteins is followed by an LHN-L motif (amino acids 128-133) which is conserved in UGT1A1, UGT1A3, UGT1A4 and UGT1A5. An F-D-KLV amino acid motif is conserved at amino acid position 128-134 in UGT1A7, UGT1A8, UGT1A9 and UGT1A10. Analysis of the N129K and R131K polymorphisms showed that these exchanges do not occur independently of one another. Homozygous UGT1A7*3 alleles were identifiable, showing that the N129K/R131K mutations can be identified on a single allele which occurs independently of UGT1A7*2. The N129/R131K polymorphism of UGT1A7 corresponds to the wild-type sequence of UGT1A9 at these codons, further suggesting that the 3 base pair mutations are inherited as a haplotype.

[0016] The third allele, UGT1A7*4, combines all the changes observed in UGT1A7*2 and UGT1A7*3. Individuals with homozygous mutations at codons 11, 129, 131 and 208 have been found, indicating that all 4 polymorphisms may occur on a single allele. It is noteworthy that there was single heterozygous or homozygous occurrence of the polymorphisms at codons 11, 129, 131 or 208. This observation also supports, as indicated above, the assumption that the polymorphisms at codons 11 and 208 and likewise at codon 129 with codon 131 are inherited as haplotypes.

[0017] On the basis of these results, the object is achieved according to the invention by a method for predicting the potential risk of carcinomas or inflammatory bowel diseases and/or for diagnosing carcinomas or inflammatory bowel diseases, in which a DNA sample from a person to be

investigated is tested for the presence of polymorphic UGT1A/alleles which include mutations at codons 11, 129, 131 and/or 208.

[0018] In particular, moreover, a positive result from a mutation at codon 11 and codon 208 is regarded as a positive sign of sensitivity for carcinomas, specifically for colorectal carcinoma or colonic carcinoma (CRC). Possible tests therefor are indicated below. A further development of the invention provides an additional check that the mutations regarded as positive are a W208R exchange and a silent mutation of codon 11 from CCC to CCA.

[0019] A positive result for mutation at all four codons 11, 129/131 and 208 is moreover regarded according to this invention as an indicator of a sensitivity for an inflammatory bowel disease, in particular the diseases comprised by the term inflammatory bowel disease (IBD) and ulcerative colitis, and a carcinoma, specifically CRC. A check that the mutations regarded as positive are N129K, R131K and W208R, and the silent mutation of CCC to CCA at codon 11, can be provided in relevant tests.

[0020] The identified association of UGT1A7 polymorphisms and colorectal carcinoma and inflammatory bowel diseases, which in turn form a risk factor for colon carcinogenesis, make [sic] it possible to use the markers described herein for identifying patients at risk of tumors. Since the detoxification enzymes for which polymorphisms were detected here carry out a global function in the human body, it is probable that relevance goes beyond the colorectal carcinoma described herein to include other cancers of the gastrointestinal tract, of the respiratory system, of blood formation and of the sex organs and glands.

[0021] One advantage of the invention is that patients can be assigned at an early date to appropriate risk groups and referred for preventive treatment, leading to an improvement in early detection.

[0022] A further advantage of the invention is that the method can also be used for diagnosis. One improvement in diagnosis is a possibility of obtaining even more quickly a clear result on the condition of the patient and accordingly of giving treatment earlier and in a more targeted manner. Early detection and immediate therapy are of crucial importance in particular for carcinomas. CRC can if detected early be prevented or cured in most cases.

[0023] Genomic DNA from lymphocytes is preferably used for an investigation. The genomic DNA can be isolated from a sample of the subject's blood by methods of column chromatography and chemical processing. This genomic DNA, which represents the genotype of the person to be investigated, forms the basis for all further genetic test strategies.

[0024] There is firstly a general provision for the DNA obtained from the patient to be amplified by the polymerase chain reaction.

[0025] A possibility for this in a first embodiment of the invention is to carry out a PCR amplification of the complete exon 1 of the UGT1A7 gene (about 855 base pairs), followed by a sequence analysis. The sequence found can be analyzed in a suitable way, i.e. be compared with the sequence of the wild type and of the UGT1A7 alleles UGT1A7*2, UGT1A7*3 and UGT1A7*4 which are relevant to this method.

[0026] In another embodiment, the only cDNA fragments to be amplified are those intended to provide information about the mutations which are sought. The DNA sample is therefore preferably treated with specific primer pairs, of which in each case one binds upstream and one binds downstream of the relevant mutated DNA regions around codons 11, 129/131 or 208, and cDNA fragments which match corresponding fragments of the wild-type allele or of the polymorphic alleles are generated in a polymerase chain reaction, and the presence of mutations at codons 11, 129, 131 and/or 208 is subsequently detected with the aid of sequencing and/or hybridization techniques.

[0027] The primer pairs used are preferably those which bind in each case approximately 50 base pairs upstream and downstream of codon 11, of codon 129/131 and of codon 208, resulting in cDNA fragments about 100 to 150 base pairs in size.

[0028] Detection of the polymorphisms with sequencing and/or hybridization methods can take place in various ways. The skilled worker is in principle free to select suitable methods from those known to him.

[0029] The polymorphisms which are sought can be investigated *inter alia* by single strand conformation polymorphism analysis (SSCP) by detecting the complete recombination of a patient's cDNA fragment obtained by the preceding PCR with a cDNA fragment, corresponding in each case, of the wild-type allele to give double strands. The underlying principle is based on the fact that an altered DNA sequence is unable to form completely congruent double strands with the wild-type sequence. This can be revealed by gel electrophoresis for the screening investigation and indicates the existence of a mutation. In a modification of this method, the polymorphic cDNA fragments can be detected by temperature gradient gel electrophoresis (TGGE).

[0030] The presence or absence of mutations would preferably be checked by sequencing the relevant cDNA by means of automated fluorescent dye sequencing.

[0031] A further possibility for detection is to reveal the presence of a polymorphism in a purely PCR-based strategy (amplification refractory mutation system) through specific primer sequences which bind at their 3' terminus to the base exchange of the polymorphism. At the same time, this method can be modified to discriminate between heterozygous and homozygous carriers. One exemplary embodiment of this method is indicated as automatable test kit system as example hereinafter. The primers used are oligonucleotides which bind on the one hand the wild-type sequence and on the other hand the polymorphic sequence, and corresponding reverse primers (antisense primers). Using these, a PCR product is always obtained only if the wild type encoded on the primer or the primer-encoded polymorphic allele is present. The controls employed are defined cDNAs comprising defined polymorphisms. The method can be automated for example by using the quantitative Taqman PCR.

[0032] The PCR conjugates can be detected in a conventional way using fluorescent dyes or enzyme markers. Sufficiently suitable techniques are available for the skilled worker to select for this purpose, and there is no need to give a special description thereof here.

[0033] The invention also encompasses test kits or test arrangements for carrying out the method of the invention.

[0034] The test arrangement belonging to this invention comprises in principle at least the genetic detection reagents

necessary for the method, namely the necessary primers or cDNAs, on a stationary support in an arrangement or sequence prepared for reading the result, there being provision for the prepared and additionally labeled DNA samples of the individual for investigation to be brought into contact with the test arrangement, to bind to one of the detection reagents and to be detected with the aid of the marker at this site. It is possible in this connection for the subject's DNA to be labeled with fluorescent dyes or radioactive isotopes.

[0035] For a combined analysis of a plurality of UGT1A7 polymorphisms, which is expedient in every case, the various detection reagents necessary for this purpose, i.e. for example the primers for the various codons 11, 129/131 and 208 or the DNA fragments of wild-type and UGT1A7*2/3/4 alleles can, depending on the chosen strategy, be immobilized together on a test arrangement, e.g. on a gene chip.

[0036] In order to facilitate reading of the results, it is advantageous for an inscription or an imprint assigned to the detection reagents to be arranged on the stationary support.

[0037] Combined analysis of all UGT1A7 polymorphisms is possible by a gene chip technique. For example, the oligonucleotide sequence of the polymorphisms and of the wild type can be immobilized in the stationary phase. Corresponding oligonucleotides from the subject's DNA are amplified by PCR and labeled by fluorescent dyes, radioactive isotopes or other conjugates. The subsequent hybridization reaction identifies the corresponding oligonucleotide on the gene array and determines the existence of a polymorphism. Various embodiments of an array-based test strategy are possible and can be adapted to this method by the skilled worker.

[0038] According to a further aspect of the invention, the polymorphic UGT1A7*2, UGT1A7*3 and UGT1A7*4 genes can be used to prepare relevant UGT isotypes. The identified polymorphisms can be expressed as recombinant proteins and be employed for example for metabolic characterization of tumor therapeutic agents in order to make a prediction of the metabolism of these substances possible. Likewise, all polymorphisms of UGT1A7 can be used to investigate the metabolism of potentially mutagenic or carcinogenic substances with the aim of making predictions about their toxicity or carcinogenic potency. For this purpose, the polymorphic alleles are expressed in heterologous expression systems, in bacteria, eukaryotic cell cultures and used subsequently as enzyme preparation for catalytic analyses. Test kits comprising a series of different polymorphism proteins makes simple *in vitro* analysis possible in this case.

[0039] The identification of susceptibility markers for colonic carcinoma and inflammatory bowel diseases makes it possible in principle to use them for gene therapy strategies. Possible uses in gene therapy in principle are currently known to the skilled worker, so that various implementation possibilities are open. The use of UGT1A7 alone or in combination with other homologous members of this enzyme family in tumor therapy is an application of the defined genetic association.

[0040] The invention therefore also encompasses the use of UGT1A7 genes or gene fragments for preparing a medicament for gene therapy treatment in the case of UGT1A7 polymorphisms of the UGT1A7*2, UGT1A7*3 and UGT1A7*4 type.

[0041] In a further development of the invention, therefore, the oral or parenteral use of the recombinant protein of genes of the UGT1A family in tumor therapy or in tumor

prevention in patients at risk is provided. The vectors which can be used for transporting UGT1A7 cDNA or homologous UGT1A cDNAs into human cells and tumor cells are those known in principle to the skilled worker and selectable by him using his expert knowledge. This leads to expression of the protein in the target tissue including embryonic or fetal tissue, where the UGT1A7 medicament for gene therapy can be used to modify precursor cells.

[0042] The UGT1A cDNA can be transported for example by parenteral vectors having organ-specific promoters, by

[0046] On use of these primers in automated quantitative PCR, conjugates with fluorescent dyes, enzymatic markers or other labeling systems for DNA are prepared and used for this diagnosis. The labeling agents are generally known and are not specially described here.

[0047] Results specifically detecting the presence of homozygous and heterozygous mutations are found with the aid of this test system. The results are depicted in table 2.

TABLE 2

Codon	Primer	Controls		Results				
		UGT1A7*1	UGT1A7*4	UGT1A7*2	UGT1A7*3	UGT1A7*1	UGT1A7*3	UGT1A7*4
11	A	-	+	+	-	+	+	+
	B	+	-	-	+	+	+	+
129	A	-	+	-	+	-	+	+
	B	+	-	+	-	+	+	-
131	A	-	+	-	+	-	+	+
	B	+	-	+	-	+	+	-
208	A	-	+	+	-	+	+	+
	B	+	-	-	+	+	+	+

intramuscular injection of UGT1A cDNA or by liposome-packaged cDNA through the gastrointestinal tract. The skilled worker is able to pick out suitable delivery systems with the aid of his expert knowledge.

[0043] The invention is explained in more detail below by means of examples and figures.

EXAMPLE 1

Test Example

[0044] Oligonucleotides (primers) which bind on the one hand the wild-type sequence and on the other hand the polymorphic sequences, and corresponding reverse primers (antisense primers) are used. A PCR product results with them only if the wild type encoded on the primer or the primer-encoded polymorphic allele is present. The principle is depicted in FIG. 1. Defined cDNAs comprising the defined polymorphisms are employed as controls.

[0045] The primers used are depicted in table 1:

TABLE 1

Polymorphism	Type of primer	Sequence	Product size
Codon 11	A	gggtggactggcctcctcca	269 bp
	B	gggtggactggcctcctccc	
Codon 129	reverse	ggcaaaaacctgaactccc	264 bp
	A	ttttttcaattgcaggagttgtttaag	
Codon 131	B	ttttttcaattgcaggagttgtttaat	271 bp
	A	aaattgcaggagttgtttaaggacaa	
Codon 208	B	aaattgcaggagttgtttaatgaccg	252 bp
	reverse	ttctaagacattttgaaaaaataggg	
Codon 208	A	gacgccatgacttcaaggagagatc	252 bp
	B	gacgccatgacttcaaggagagatc	
	reverse	tggcttccctgatgacagttgatac	

[0048] The use of these analyses for all three polymorphic gene loci 11, 129/131 and 208 is therefore of great importance because the presence of the UGT1A7*2 allele (codon 11 and 208) and of the UGT1A7*4 allele (codon 11, 129, 131, 208) are [sic] associated with colorectal carcinoma, but only the presence of the UGT1A7*4 allele are [sic] associated with inflammatory bowel diseases. The genetic analysis therefore expediently consists of combination of the analyses at the identified loci in the sense of a test kit.

Test Series on 111 Subjects

[0049] The samples were collected from 111 patients of the Department of Gastroenterology and Hepatology of Hannover Medical School in 1999.

[0050] The following groups were investigated:

[0051] a) the control group (n=54, 44.32±15.65 years, 31 male/23 female) was defined as patients who had no finding of previous cancers, precluded by abdominal ultrasound and endoscopy (upper and lower).

[0052] b) colorectal cancer patients (n=26, 62.15±11.25 years, 15 male/11 female) suffered from colorectal carcinoma, as confirmed by exclusion according to the Amsterdam II criteria.

[0053] c) the examples of inflammatory bowel diseases (IBD) (n=31, 36.74±11.79 years, 14 male/16 female) comprised patients with ulcerative colitis (n=14, 35.07±13.86 years, 9 male/5 female) and Crohn's disease (n=17, 38.12±10.00 years 5 male/12 female). The diagnoses of ulcerative colitis and of Crohn's disease were based on findings of upper and lower endoscopy and were confirmed by histology of intestinal biopsies.

[0054] Genomic DNA: The genomic DNA was prepared from whole blood samples using the QuiaAmp® system in

accordance with the manufacturer's recommendations (Quiagen, Hilden, Germany). Concentration data were determined by spectrophotometry at 260 and 280 nm, and the samples were stored in 10 mM Tris/EDTA buffer (pH 8.0) at 4° C. until investigated further.

[0055] Analysis of the UGT1A7 exon 1 sequence: The UGT1A7 exon 1 sequence was amplified by the polymerase chain reaction. The forward primer was from base pair 61 to 38 upstream of the ATG start codon (GenBank access number U39570) of UGT1A7 (5'-gcggtctgagccactactatattataggagct-3'). The reverse primer was between base pair 855 and 829 (GenBank access number U89507) of the UGT1A7 exon 1 sequence (5'-gcgatcatccataggaactgcttcctgatgaca-3'). Location of the upstream primer outside the open reading frame was necessary in order to achieve specificity of amplification, because the exon 1 sequence of UGT1A7-10 has homologies of 93%. A product 916 base pairs in size was amplified in a volume of 100 μ l comprising 10 mM KCl, 20 mM Tris-HCl (pH 8.8), 10 mM ammonium sulfate, 2 mM magnesium sulfate, 1% Triton X-100, 0.2 mM each dNTP, 20 ng of genomic DNA, 2 μ M primer and 5 units of VENT (exo) DNA polymerase (NEB, Beverly, Mass.). After a hot start at 94° C. for 3 minutes, 30 cycles of 94° C. for 3 seconds, 57° C. for 30 seconds and 72° C. for 30 seconds were run on a Perkin Elmer Gene Amp PCR 2400 system. The products were visualized in a 2% agarose gel electrophoresis and purified using QuiaQuick® columns according to the manufacturer's statements (Quiagen, Hilden, Germany). The sequences of the PCR products were determined on both strands by automated fluorescent dye sequencing (MWG-Biotech, Ebersbach, Germany). The sequence data were analyzed using the PC gene software package (Oxford Molecular, Campbell, Calif., USA). The statistical analysis was calculated using the Kruskal-Wallis test.

[0056] GenBank accession numbers: UGT1A7*1: U89507; UGT1A7*2: AF2969226; UGT1A7*3: AF292627, UGT1A7*4 AF292627

[0057] The invention is explained in more detail by means of figures below, these specifically showing:

[0058] FIG. 1 graphical representation of the UGT1A7 polymorphism diagnostic method;

[0059] FIG. 2 association of UGT1A7*2 and UGT1A7*4 with NC, CRC and IBD;

[0060] FIG. 3 3 polymorphic sites of the human UGT1A7 exon 1 sequence;

[0061] FIG. 4 comparison of the regions around codon 11, 129, 131 and 208;

[0062] FIG. 5 associations between the W208R mutation and colorectal carcinoma and inflammatory bowel diseases.

[0063] FIG. 1 shows a UGT1A7 polymorphism diagnostic method in which the region around the identified polymorphisms is amplified by polymerase chain reaction (PCR). Specific primer pairs which bind about 50 base pairs upstream and downstream of codon 11, of codon 129/131 and of codon 208 are used for this. This PCR generates complementary deoxyribonucleic acid fragments (cDNAs) which are 100 to 150 bp in size and which correspond to the wild-type allele or carry the identified polymorphisms.

[0064] FIG. 2 shows the analysis of the prevalence of the polymorphic UGT1A7 alleles as defined in FIG. 3. This analysis, which is based on the defined polymorphic UGT1A7 alleles, shows a significant association of UGT1A7*4 and UGT1A7*2 with colorectal cancer (CRC) and the low occurrence of these alleles in the normal control group (NC). In contrast thereto, inflammatory bowel diseases (IBD) are associated only with the presence of UGT1A7*4. No significant association with the UGT1A7*3 allele was found either for CRC or for IBD. The association of linkages of the heterozygous UGT1A7*3/UGT1A7*4 individuals with CRC and IBD was, however, significant. Statistically significant comparisons are indicated by *. UGT1A7*2 is defined as a combination of CCA at codon 11 and the W208R mutation. This is based on the observation that the two mutations are always identified in combination and not individually heterozygous or homozygous either at codon 11 or codon 208. UGT1A7*2 was not detected as homozygous allele. UGT1A7*3 is defined only by the combination of N129K and R131K. These patents [sic] were either homozygous or heterozygous both for N129K and R131K. The heterozygous UGT1A7*4 may alternatively represent the linkage heterozygous UGT1A7*2/UGT1A7*3.

[0065] FIG. 3 shows a graphical representation of the 3 polymorphic sites identified in the human UGT1A7 exon 1 sequence. The mutation CCC/A at codon 11 is silent. The mutations at codon 129, 131 and 208 lead to amino acid substitutions. Sequence analysis indicated the presence of 3 different polymorphisms which were assigned to UGT1A7*2, UGT1A7*3 and UGT1A7*4. The exchange at codon 11 and 208 (UGT1A7*2) and the polymorphisms at codon 129 and 131 (UGT1A7*3) did not occur singly.

[0066] FIG. 4 shows that, compared with the region around codon 11, the identified polymorphisms at position 129/131 and position 208 occur in highly conserved sequence regions. The N129K/R131K polymorphism represents the wild-type sequence of UGT1A9 in this position. W208R represents the wild-type sequence of UGT1A8 and UGT1A9 at this position.

[0067] FIG. 5 shows that the association of W208R mutations with colorectal carcinomas and inflammatory bowel diseases is highly statistically significant. Whereas mutations at codon 208 (W208R) occur in only 22% of normal controls, they are present in 73% of patients with colorectal carcinoma and in 61% of patients with inflammatory bowel diseases. A significant association with colorectal carcinoma (19% versus 7%) but not with inflammatory bowel disease was found for the UGT1A7*2 allele. The UGT1A7*4 allele was significantly associated with colorectal carcinoma and also with inflammatory bowel diseases, whereas no significant association was found for UGT1A7*3 (low significance level). The statistical significance for the presence of W208R mutations in CRC patients (73%) compared with normal controls (22%) was very high ($p < 0.0001$). This included heterozygous (50% versus 15%, $p < 0.001$) and homozygous (23% versus 7%, $p < 0.02$) W208R mutations. The association was less but likewise significant for IBD patients, who showed 61% W208R mutations ($p < 0.0001$), in 35% heterozygous and 26% homozygous individuals. An additional point to take into account concerning the 22% occurrence of W208R mutations in normal controls is that a tumor-free state at the time the blood sample was taken does not preclude later development of CRC.

[0068] The * shows significant level compared with the normal controls (Kruskal Wallis test). The meanings hereinafter are: n.s.—not significant, NC—normal controls; CRC—colorectal carcinoma and IBD—inflammatory bowel diseases.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 57

<210> SEQ ID NO 1

<211> LENGTH: 33

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 1

gcggtcgag ccacttacta tattatagga gct 33

<210> SEQ ID NO 2

<211> LENGTH: 36

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 2

gcggatatcc ataggcactg gctttcctg atgaca 36

<210> SEQ ID NO 3

<211> LENGTH: 21

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 3

gggtggactg gctccttcc a 21

<210> SEQ ID NO 4

<211> LENGTH: 21

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 4

gggtggactg gctccttcc c 21

<210> SEQ ID NO 5

<211> LENGTH: 21

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 5

ggcaaaaacc atgaactccc g 21

<210> SEQ ID NO 6

<211> LENGTH: 30

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 6

ttttttcaa attgcaggag tttgtttaag 30

<210> SEQ ID NO 7

<211> LENGTH: 30

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

-continued

<400> SEQUENCE: 7
 ttttttcaa attgcaggag tttgtttaat 30

<210> SEQ ID NO 8
 <211> LENGTH: 27
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 8
 aaattgcagg agtttgttta aggacaa 27

<210> SEQ ID NO 9
 <211> LENGTH: 27
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 9
 aaattgcagg agtttgttta atgaccg 27

<210> SEQ ID NO 10
 <211> LENGTH: 27
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 10
 ttctaagaca ttttgaaaa aataggg 27

<210> SEQ ID NO 11
 <211> LENGTH: 28
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 11
 gacgcatga ctttcaagga gagagtac 28

<210> SEQ ID NO 12
 <211> LENGTH: 28
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 12
 gacgcatga ctttcaagga gagagtat 28

<210> SEQ ID NO 13
 <211> LENGTH: 27
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 13
 tggctttccc tgatgacagt tgatacc 27

<210> SEQ ID NO 14
 <211> LENGTH: 530
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 14
 Met Ala Arg Ala Gly Trp Thr Gly Leu Leu Pro Leu Tyr Val Cys Leu
 1 5 10 15

-continued

Leu Leu Thr Cys Gly Phe Ala Lys Ala Gly Lys Leu Leu Val Val Pro
 20 25 30
 Met Asp Gly Ser His Trp Phe Thr Met Gln Ser Val Val Glu Lys Leu
 35 40 45
 Ile Leu Arg Gly His Glu Val Val Val Val Met Pro Glu Val Ser Trp
 50 55 60
 Gln Leu Gly Arg Ser Leu Asn Cys Thr Val Lys Thr Tyr Ser Thr Ser
 65 70 75 80
 Tyr Thr Leu Glu Asp Gln Asp Arg Glu Phe Met Val Phe Ala Asp Ala
 85 90 95
 Arg Trp Thr Ala Pro Leu Arg Ser Ala Phe Ser Leu Leu Thr Ser Ser
 100 105 110
 Ser Asn Gly Ile Phe Asp Leu Phe Phe Ser Asn Cys Arg Ser Leu Phe
 115 120 125
 Asn Asp Arg Lys Leu Val Glu Tyr Leu Lys Glu Ser Cys Phe Asp Ala
 130 135 140
 Val Phe Leu Asp Pro Phe Asp Ala Cys Gly Leu Ile Val Ala Lys Tyr
 145 150 155 160
 Phe Ser Leu Pro Ser Val Val Phe Ala Arg Gly Ile Phe Cys His Tyr
 165 170 175
 Leu Glu Glu Gly Ala Gln Cys Pro Ala Pro Leu Ser Tyr Val Pro Arg
 180 185 190
 Leu Leu Leu Gly Phe Ser Asp Ala Met Thr Phe Lys Glu Arg Val Trp
 195 200 205
 Asn His Ile Met His Leu Glu Glu His Leu Phe Cys Pro Tyr Phe Phe
 210 215 220
 Lys Asn Val Leu Glu Ile Ala Ser Glu Ile Leu Gln Thr Pro Val Thr
 225 230 235 240
 Ala Tyr Asp Leu Tyr Ser His Thr Ser Ile Trp Leu Leu Arg Thr Asp
 245 250 255
 Phe Val Leu Glu Tyr Pro Lys Pro Val Met Pro Asn Met Ile Phe Ile
 260 265 270
 Gly Gly Ile Asn Cys His Gln Gly Lys Pro Val Pro Met Glu Phe Glu
 275 280 285
 Ala Tyr Ile Asn Ala Ser Gly Glu His Gly Ile Val Val Phe Ser Leu
 290 295 300
 Gly Ser Met Val Ser Glu Ile Pro Glu Lys Lys Ala Met Ala Ile Ala
 305 310 315 320
 Asp Ala Leu Gly Lys Ile Pro Gln Thr Val Leu Trp Arg Tyr Thr Gly
 325 330 335
 Thr Arg Pro Ser Asn Leu Ala Asn Asn Thr Ile Leu Val Lys Trp Leu
 340 345 350
 Pro Gln Asn Asp Leu Leu Gly His Pro Met Thr Arg Ala Phe Ile Thr
 355 360 365
 His Ala Gly Ser His Gly Val Tyr Glu Ser Ile Cys Asn Gly Val Pro
 370 375 380
 Met Val Met Met Pro Leu Phe Gly Asp Gln Met Asp Asn Ala Lys Arg
 385 390 395 400
 Met Glu Thr Lys Gly Ala Gly Val Thr Leu Asn Ala Leu Glu Met Thr
 405 410 415
 Ser Glu Asp Leu Glu Asn Ala Leu Lys Ala Val Ile Asn Asp Lys Ser

-continued

Ala Tyr Asp Leu Tyr Ser His Thr Ser Ile Trp Leu Leu Arg Thr Asp
 245 250 255

 Phe Val Leu Glu Tyr Pro Lys Pro Val Met Pro Asn Met Ile Phe Ile
 260 265 270

 Gly Gly Ile Asn Cys His Gln Gly Lys Pro Val Pro Met Glu Phe Glu
 275 280 285

 Ala Tyr Ile Asn Ala Ser Gly Glu His Gly Ile Val Val Phe Ser Leu
 290 295 300

 Gly Ser Met Val Ser Glu Ile Pro Glu Lys Lys Ala Met Ala Ile Ala
 305 310 315 320

 Asp Ala Leu Gly Lys Ile Pro Gln Thr Val Leu Trp Arg Tyr Thr Gly
 325 330 335

 Thr Arg Pro Ser Asn Leu Ala Asn Asn Thr Ile Leu Val Lys Trp Leu
 340 345 350

 Pro Gln Asn Asp Leu Leu Gly His Pro Met Thr Arg Ala Phe Ile Thr
 355 360 365

 His Ala Gly Ser His Gly Val Tyr Glu Ser Ile Cys Asn Gly Val Pro
 370 375 380

 Met Val Met Met Pro Leu Phe Gly Asp Gln Met Asp Asn Ala Lys Arg
 385 390 395 400

 Met Glu Thr Lys Gly Ala Gly Val Thr Leu Asn Ala Leu Glu Met Thr
 405 410 415

 Ser Glu Asp Leu Glu Asn Ala Leu Lys Ala Val Ile Asn Asp Lys Ser
 420 425 430

 Phe Lys Glu Asn Ile Met Arg Leu Ser Ser Leu His Lys Asp Arg Pro
 435 440 445

 Val Glu Pro Leu Asp Leu Ala Val Phe Trp Val Glu Phe Val Met Arg
 450 455 460

 His Lys Gly Ala Pro His Leu Arg Pro Ala Ala His Asp Leu Thr Trp
 465 470 475 480

 Tyr Gln Tyr His Ser Leu Asp Val Ile Gly Phe Leu Leu Ala Val Val
 485 490 495

 Leu Thr Val Ala Phe Ile Thr Phe Lys Cys Cys Ala Tyr Gly Tyr Arg
 500 505 510

 Lys Cys Leu Gly Lys Lys Gly Arg Val Lys Lys Ala His Lys Ser Lys
 515 520 525

 Thr His
 530

<210> SEQ ID NO 16

<211> LENGTH: 530

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 16

Met Ala Arg Ala Gly Trp Thr Gly Leu Leu Pro Leu Tyr Val Cys Leu
 1 5 10 15

 Leu Leu Thr Cys Gly Phe Ala Lys Ala Gly Lys Leu Leu Val Val Pro
 20 25 30

 Met Asp Gly Ser His Trp Phe Thr Met Gln Ser Val Val Glu Lys Leu
 35 40 45

 Ile Leu Arg Gly His Glu Val Val Val Val Met Pro Glu Val Ser Trp
 50 55 60

-continued

Gln Leu Gly Arg Ser Leu Asn Cys Thr Val Lys Thr Tyr Ser Thr Ser
 65 70 75 80
 Tyr Thr Leu Glu Asp Gln Asp Arg Glu Phe Met Val Phe Ala Asp Ala
 85 90 95
 Arg Trp Thr Ala Pro Leu Arg Ser Ala Phe Ser Leu Leu Thr Ser Ser
 100 105 110
 Ser Asn Gly Ile Phe Asp Leu Phe Phe Ser Asn Cys Arg Ser Leu Phe
 115 120 125
 Lys Asp Lys Lys Leu Val Glu Tyr Leu Lys Glu Ser Cys Phe Asp Ala
 130 135 140
 Val Phe Leu Asp Pro Phe Asp Ala Cys Gly Leu Ile Val Ala Lys Tyr
 145 150 155 160
 Phe Ser Leu Pro Ser Val Val Phe Ala Arg Gly Ile Phe Cys His Tyr
 165 170 175
 Leu Glu Glu Gly Ala Gln Cys Pro Ala Pro Leu Ser Tyr Val Pro Arg
 180 185 190
 Leu Leu Leu Gly Phe Ser Asp Ala Met Thr Phe Lys Glu Arg Val Trp
 195 200 205
 Asn His Ile Met His Leu Glu Glu His Leu Phe Cys Pro Tyr Phe Phe
 210 215 220
 Lys Asn Val Leu Glu Ile Ala Ser Glu Ile Leu Gln Thr Pro Val Thr
 225 230 235 240
 Ala Tyr Asp Leu Tyr Ser His Thr Ser Ile Trp Leu Leu Arg Thr Asp
 245 250 255
 Phe Val Leu Glu Tyr Pro Lys Pro Val Met Pro Asn Met Ile Phe Ile
 260 265 270
 Gly Gly Ile Asn Cys His Gln Gly Lys Pro Val Pro Met Glu Phe Glu
 275 280 285
 Ala Tyr Ile Asn Ala Ser Gly Glu His Gly Ile Val Val Phe Ser Leu
 290 295 300
 Gly Ser Met Val Ser Glu Ile Pro Glu Lys Lys Ala Met Ala Ile Ala
 305 310 315 320
 Asp Ala Leu Gly Lys Ile Pro Gln Thr Val Leu Trp Arg Tyr Thr Gly
 325 330 335
 Thr Arg Pro Ser Asn Leu Ala Asn Asn Thr Ile Leu Val Lys Trp Leu
 340 345 350
 Pro Gln Asn Asp Leu Leu Gly His Pro Met Thr Arg Ala Phe Ile Thr
 355 360 365
 His Ala Gly Ser His Gly Val Tyr Glu Ser Ile Cys Asn Gly Val Pro
 370 375 380
 Met Val Met Met Pro Leu Phe Gly Asp Gln Met Asp Asn Ala Lys Arg
 385 390 395 400
 Met Glu Thr Lys Gly Ala Gly Val Thr Leu Asn Ala Leu Glu Met Thr
 405 410 415
 Ser Glu Asp Leu Glu Asn Ala Leu Lys Ala Val Ile Asn Asp Lys Ser
 420 425 430
 Phe Lys Glu Asn Ile Met Arg Leu Ser Ser Leu His Lys Asp Arg Pro
 435 440 445
 Val Glu Pro Leu Asp Leu Ala Val Phe Trp Val Glu Phe Val Met Arg
 450 455 460

-continued

Ala Tyr Ile Asn Ala Ser Gly Glu His Gly Ile Val Val Phe Ser Leu
 290 295 300

Gly Ser Met Val Ser Glu Ile Pro Glu Lys Lys Ala Met Ala Ile Ala
 305 310 315 320

Asp Ala Leu Gly Lys Ile Pro Gln Thr Val Leu Trp Arg Tyr Thr Gly
 325 330 335

Thr Arg Pro Ser Asn Leu Ala Asn Asn Thr Ile Leu Val Lys Trp Leu
 340 345 350

Pro Gln Asn Asp Leu Leu Gly His Pro Met Thr Arg Ala Phe Ile Thr
 355 360 365

His Ala Gly Ser His Gly Val Tyr Glu Ser Ile Cys Asn Gly Val Pro
 370 375 380

Met Val Met Met Pro Leu Phe Gly Asp Gln Met Asp Asn Ala Lys Arg
 385 390 395 400

Met Glu Thr Lys Gly Ala Gly Val Thr Leu Asn Ala Leu Glu Met Thr
 405 410 415

Ser Glu Asp Leu Glu Asn Ala Leu Lys Ala Val Ile Asn Asp Lys Ser
 420 425 430

Phe Lys Glu Asn Ile Met Arg Leu Ser Ser Leu His Lys Asp Arg Pro
 435 440 445

Val Glu Pro Leu Asp Leu Ala Val Phe Trp Val Glu Phe Val Met Arg
 450 455 460

His Lys Gly Ala Pro His Leu Arg Pro Ala Ala His Asp Leu Thr Trp
 465 470 475 480

Tyr Gln Tyr His Ser Leu Asp Val Ile Gly Phe Leu Leu Ala Val Val
 485 490 495

Leu Thr Val Ala Phe Ile Thr Phe Lys Cys Cys Ala Tyr Gly Tyr Arg
 500 505 510

Lys Cys Leu Gly Lys Lys Gly Arg Val Lys Lys Ala His Lys Ser Lys
 515 520 525

Thr His
 530

<210> SEQ ID NO 18
 <211> LENGTH: 1593
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 18

```

atggtctctg cagggtggac tggcctcctt ccctatatg tgtgtctact gctgacctgt    60
ggctttgccaggcaggaa gctgctggta gtgccatgg atgggagcca ctggttcacc    120
atgcagtcgg tggtggagaa actcatcctc agggggcatg aggtggctgt agtcatgccaa    180
gaggtgagtt ggcaactggg aagatcactg aattgcacag tgaagactta ctcaacctca    240
tacactctgg aggatcagga ccgggagttc atggtttttg ccgatgctcg ctggacggca    300
ccattgcgaa gtgcattttc tctattaaca agttcatcca atggatattt tgacttattt    360
ttttcaaatt gcaggagttt gtttaatgac cgaaaattag tagaatactt aaaggagagt    420
tgttttgatg cagtgtttct cgatcctttt gatgcctgtg gcttaattgt tgccaaatat    480
ttctccctcc cctctgtggt ctccgccagg ggaatatttt gccactatct tgaagaaggt    540
gcacagtgcc ctgctcctct ttcctatgct cccagacttc tcttagggtt ctacagcgcc    600
    
```

-continued

```

atgactttca aggagagagt atggaaccac atcatgcaact tggaggaaca tttatTTTTgc 660
ccctatTTTT tcaaaaatgt cttagaaata gcctctgaaa ttctccaaac cctgtcacg 720
gcatatgata tctacagcca cacatcaatt tggttgTTgc gaactgactt tgttttggag 780
tatcccaaac cctgatgcc caatatgata ttcatggtg gtatcaactg tcatcagga 840
aagccagtgc ctatggaatt tgaagcctac attaatgctt ctggagaaca tggaaTTgtg 900
gttttctctt tgggatcaat ggtctcagaa attccagaga agaaagctat ggcaattgct 960
gatgcttTgg gcaaaaatccc tcagacagtc ctgtggcggg aactggaac cggaccatcg 1020
aatcttgcga acaacacgat actgtttaag tggctacccc aaaacgatct gcttggTcac 1080
ccgatgccc gtgcctttat caccatgct ggttcccatg gtgtttatga aagcatatgc 1140
aatggcgttc ccatggtgat gatgcccttg tttggtgata agatggacaa tgcaaaagcgc 1200
atggagacta agggagctgg agtgaccctg aatgctctgg aatgacttc tgaagattta 1260
gaaaaatgctc taaaagcagt catcaatgac aaaagtttca aggagaacat catgcgcctc 1320
tccagccttc acaaggaccg cccgggtggag ccgctggacc tggccgtggt ctgggtggag 1380
tttgtgatga ggcacaaggc gcgcgccacac ctgcgccccg cagcccacga cctcacctgg 1440
taccagtacc attccttTga cgtgattggt ttctcttTgg ccgtcgtgct gacagtggcc 1500
ttcatcacct ttaaatgttg tgcttatggc taccggaaat gcttggggaa aaaagggcga 1560
gttaagaaag cccacaatc caagaccat tga 1593

```

```

<210> SEQ ID NO 19
<211> LENGTH: 1593
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 19

```

```

atggctcgtg cagggtggac tggcctcctt ccactatatg tgtgtctact gctgacctgt 60
ggctttgcc aaggcaggaa gctgctggta gtgccatgg atgggagcca ctggttcacc 120
atgcagtcgg tggTggagaa actcatcctc agggggcatg aggtggtcgt agtcatgcca 180
gaggtgagtt ggcaactggg aagatcactg aattgcacag tgaagactta ctcaacctca 240
tacactctgg aggatcagga ccgggagttc atggTTTTtg ccgatgctcg ctggacggca 300
ccattgcgaa gtgcattttc tctattaaca agttcatcca atggtatttt tgacttattt 360
ttttcaaatt gcaggagttt gtttaatgac cgaaaattag tagaatactt aaaggagagt 420
tgttttgatg cagtgtttct cgatcctttt gatgcctgtg gcttaattgt tgccaaatat 480
ttctccctcc cctctgtggt cttcgcagc ggaatatttt gccactatct tgaagaaggt 540
gcacagtgcc ctgctcctct ttctatgct cccagacttc tcttagggtt ctgagacgcc 600
atgactttca aggagagagt acggaaccac atcatgcaact tggaggaaca tttatTTTTgc 660
ccctatTTTT tcaaaaatgt cttagaaata gcctctgaaa ttctccaaac cctgtcacg 720
gcatatgata tctacagcca cacatcaatt tggttgTTgc gaactgactt tgttttggag 780
tatcccaaac cctgatgcc caatatgata ttcatggtg gtatcaactg tcatcagga 840
aagccagtgc ctatggaatt tgaagcctac attaatgctt ctggagaaca tggaaTTgtg 900
gttttctctt tgggatcaat ggtctcagaa attccagaga agaaagctat ggcaattgct 960
gatgcttTgg gcaaaaatccc tcagacagtc ctgtggcggg aactggaac cggaccatcg 1020

```

-continued

```

aatcttgca acaacacgat acttgtaag tggctacccc aaaacgatct gcttggtcac 1080
ccgatgacc gtgcctttat caccatgct ggttcccatg gtgtttatga aagcatatgc 1140
aatggcgctt ccatgggtgat gatgcccttg tttggtgatc agatggacaa tgcaaagcgc 1200
atggagacta agggagctgg agtgaccctg aatgctctgg aaatgacttc tgaagattta 1260
gaaaatgctc taaaagcagt catcaatgac aaaagtttca aggagaacat catgcccctc 1320
tccagccttc acaaggaccg cccgggtggag ccgctggacc tggccgtggt ctgggtggag 1380
tttgtgatga ggcacaaggc cgcgccacac ctgcgccccg cagcccacga cctcacctgg 1440
taccagtacc attccttga cgtgattggt ttctcttgg ccgtcgtgct gacagtggcc 1500
ttcatcacct taaatgttg tgcttatggc taccgaaat gcttggggaa aaaagggcga 1560
gttaagaaag cccacaatc caagacccat tga 1593

```

<210> SEQ ID NO 20

<211> LENGTH: 1593

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 20

```

atggctcgtg cagggtggac tggcctcctt ccctatatg tgtgtctact gctgacctgt 60
ggctttgcca aggcaggaa gctgctggta gtgccatgg atgggagcca ctggttcacc 120
atgcagtcgg tggtgagaa actcatcctc agggggcatg aggtggtcgt agtcatgcca 180
gaggtgagtt ggcaactggg aagatcactg aattgcacag tgaagactta ctcaacctca 240
tacctctgg aggatcagga cggggagttc atggtttttg ccgatgctcg ctggacggca 300
ccattgcgaa gtgcattttc tctattaaca agttcatcca atggatattt tgacttattt 360
ttttcaaatt cgaggagttt gtttaaggac aaaaaattag tagaatactt aaaggagagt 420
tgttttgatg cagtgtttct cgatcctttt gatgcctgtg gcttaattgt tgccaaatat 480
ttctccctcc cctctgtggt cttcgcagg ggaatattt gccactatct tgaagaaggt 540
gcacagtgcc ctgctcctct ttcctatgtc ccagacttc tcttagggtt ctcagacgcc 600
atgactttca aggagagagt atggaaccac atcatgcaact tggaggaaca tttattttgc 660
ccctattttt tcaaaaatgt cttagaataa gcctctgaaa ttctccaaac ccctgtcacc 720
gcatatgata tctacagcca cacatcaatt tggttgttgc gaactgactt tgttttgag 780
tatcccaaac ccgtgatgcc caatatgata ttcattggtg gtatcactct tcatcaggga 840
aagccagtgc ctatggaatt tgaagcctac attaatgctt ctggagaaca tggaaattgtg 900
gttttctctt tgggatcaat ggtctcagaa attccagaga agaaagctat ggcaattgct 960
gatgcttttg gcaaaaatcc tcagacagtc ctgtggcggg aactggaac ccgaccatcg 1020
aatcttgca acaacacgat acttgtaag tggctacccc aaaacgatct gcttggtcac 1080
ccgatgacc gtgcctttat caccatgct ggttcccatg gtgtttatga aagcatatgc 1140
aatggcgctt ccatgggtgat gatgcccttg tttggtgatc agatggacaa tgcaaagcgc 1200
atggagacta agggagctgg agtgaccctg aatgctctgg aaatgacttc tgaagattta 1260
gaaaatgctc taaaagcagt catcaatgac aaaagtttca aggagaacat catgcccctc 1320
tccagccttc acaaggaccg cccgggtggag ccgctggacc tggccgtggt ctgggtggag 1380
tttgtgatga ggcacaaggc cgcgccacac ctgcgccccg cagcccacga cctcacctgg 1440

```

-continued

```
taccagtacc attccttggg cgtgattggt ttctctttgg ccgctgtgct gacagtggcc 1500
ttcatcacct ttaaatgttg tgcttatggc taccggaaat gcttggggaa aaaagggcga 1560
gttaagaaag cccacaaatc caagacccat tga 1593
```

```
<210> SEQ ID NO 21
<211> LENGTH: 1593
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
```

```
<400> SEQUENCE: 21
```

```
atggctcgtg cagggtggac tggcctcctt ccactatatg tgtgtctact gctgacctgt 60
ggctttgccg aggcagggaa gctgctggta gtgccatgg atgggagcca ctggttcacc 120
atgcagtcgg tggtgagaaa actcatcctc agggggcatg aggtggctgt agtcatgcc 180
gaggtgagtt ggcaactggg aagatcactg aattgcacag tgaagactta ctcaacctca 240
tacactctgg aggatcagga ccgggagttc atggtttttg ccgatgctcg ctggacggca 300
ccattgcaaa gtgcattttc tctattaaca agttcatcca atggatattt tgacttattt 360
ttttcaaat gcaggagttt gttaagagc aaaaaattag tagaatactt aaaggagagt 420
tgttttgatg cagtgtttct cgatcctttt gatgcctgtg gcttaattgt tgccaaatat 480
ttctccctcc cctctgtggt cttcgcagg ggaatattt gccactatct tgaagaaggt 540
gcacagtgcc ctgctcctt ttctctatgt cccagacttc tcttaggggt ctgagcggc 600
atgactttca aggagagagt acggaaccac atcatgcact tggaggaaca tttattttgc 660
ccctattttt tcaaaaatgt ctagaata gcctctgaaa ttctocaaac ccctgtcagc 720
gcatatgatc tctacagcca cacatcaatt tggttgttc gaactgactt tgttttgag 780
tatcccaaac ccgtgatgc caatatgatc ttcattggtg gtatcaactg tcatcagga 840
aagccagtgc ctatggaatt tgaagcctac attaatgctt ctggagaaca tggaaattgtg 900
gttttctctt tgggatcaat ggtctcagaa attccagaga agaaagctat ggcaattgct 960
gatgctttgg gcaaaaatcc tcagacagtc ctgtggcggg aactggaac ccgaccatcg 1020
aatcttgca acaacacgat acttgtaag tggctacccc aaaacgatct gcttggtcac 1080
ccgatgccc gtgcctttat caccatgct ggttccatg gtgtttatga aagcatatgc 1140
aatggcgttc ccatggtgat gatgcccttg tttggtgatc agatggacaa tgcaaaagc 1200
atggagacta agggagctgg agtgaccctg aatgctctgg aatgacttc tgaagattta 1260
gaaaatgctc taaaagcagt catcaatgac aaaagtttca aggagaacat catgagcctc 1320
tccagccttc acaaggaccg cccgggtggg ccgctggacc tggccgtggt ctgggtggag 1380
tttgtgatga ggcacaaggc cgcgccacac ctgcccccg cagcccacga cctcacctgg 1440
taccagtacc attccttggg cgtgattggt ttctctttgg ccgctgtgct gacagtggcc 1500
ttcatcacct ttaaatgttg tgcttatggc taccggaaat gcttggggaa aaaagggcga 1560
gttaagaaag cccacaaatc caagacccat tga 1593
```

```
<210> SEQ ID NO 22
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
```

-continued

<222> LOCATION: (1)..(15)
<223> OTHER INFORMATION: DNA encoding residues 128-132 of UGT1A7*1 as
shown in Figure 3

<400> SEQUENCE: 22

tttaatgacc gaaaa 15

<210> SEQ ID NO 23
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(5)
<223> OTHER INFORMATION: amino acid residues 128-132 of UGT1A7*1 as
shown in Figure 3

<400> SEQUENCE: 23

Phe Asn Asp Arg Lys
1 5

<210> SEQ ID NO 24
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(15)
<223> OTHER INFORMATION: DNA encoding residues 128-132 of UGT1A7*2 as
shown in Figure 3

<400> SEQUENCE: 24

tttaatgacc gaaaa 15

<210> SEQ ID NO 25
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(5)
<223> OTHER INFORMATION: amino acid residues 128-132 of UGT1A7*2 as
shown in Figure 3

<400> SEQUENCE: 25

Phe Asn Asp Arg Lys
1 5

<210> SEQ ID NO 26
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(15)
<223> OTHER INFORMATION: DNA encoding residues 128-132 of UGT1A7*3 as
shown in Figure 3

<400> SEQUENCE: 26

tttaaggaca aaaaa 15

<210> SEQ ID NO 27
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:

-continued

<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(5)
<223> OTHER INFORMATION: amino acid residues 128-132 of UGT1A7*3 as
shown in Figure 3

<400> SEQUENCE: 27

Phe Lys Asp Lys Lys
1 5

<210> SEQ ID NO 28
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(15)
<223> OTHER INFORMATION: DNA encoding residues 128-132 of UGT1A7*4 as
shown in Figure 3

<400> SEQUENCE: 28

tttaaggaca aaaaa

15

<210> SEQ ID NO 29
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(5)
<223> OTHER INFORMATION: amino acid residues 128-132 of UGT1A7*4 as
shown in Figure 3

<400> SEQUENCE: 29

Phe Lys Asp Lys Lys
1 5

<210> SEQ ID NO 30
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(4)
<223> OTHER INFORMATION: amino acid residues 7-10 of UGT1A1 as
shown in Figure 4

<400> SEQUENCE: 30

Gly Gly Arg Pro
1

<210> SEQ ID NO 31
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(4)
<223> OTHER INFORMATION: amino acid residues 12-15 of UGT1A3 as
shown in Figure 4

<400> SEQUENCE: 31

Leu Val Leu Gly
1

<210> SEQ ID NO 32
<211> LENGTH: 8

-continued

<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: amino acid residues 127-134 of UGT1A1 as
shown in Figure 4

<400> SEQUENCE: 32

Leu Leu His Asn Lys Glu Leu Met
1 5

<210> SEQ ID NO 33
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 205-1213of UGT1A1 as
shown in Figure 4

<400> SEQUENCE: 33

Gln Arg Val Lys Asn Met Leu Ile Ala
1 5

<210> SEQ ID NO 34
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 7-15 of UGT1A3 as
shown in Figure 4

<400> SEQUENCE: 34

Val Pro Leu Pro Trp Leu Ala Thr Gly
1 5

<210> SEQ ID NO 35
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: amino acid residues 127-134 of UGT1A3 as
shown in Figure 4

<400> SEQUENCE: 35

Leu Leu His Asn Glu Ala Leu Ile
1 5

<210> SEQ ID NO 36
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 205-1213f UGT1A3 as
shown in Figure 4

<400> SEQUENCE: 36

Gln Arg Val Lys Asn Met Leu Tyr Pro
1 5

-continued

<210> SEQ ID NO 37
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 7-15 of UGT1A4 as
shown in Figure 4

<400> SEQUENCE: 37

Val Pro Leu Pro Gln Leu Ala Thr Gly
1 5

<210> SEQ ID NO 38
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: amino acid residues 127-134 of UGT1A4 as
shown in Figure 4

<400> SEQUENCE: 38

Leu Leu His Asn Glu Ala Leu Ile
1 5

<210> SEQ ID NO 39
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 205-213 of UGT1A4 as
shown in Figure 4

<400> SEQUENCE: 39

Gln Arg Val Lys Asn Met Leu Tyr Pro
1 5

<210> SEQ ID NO 40
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 7-15 of UGT1A5 as
shown in Figure 4

<400> SEQUENCE: 40

Val Pro Leu Pro Arg Leu Ala Thr Gly
1 5

<210> SEQ ID NO 41
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: amino acid residues 127-134 of UGT1A5 as
shown in Figure 4

-continued

<400> SEQUENCE: 41

Leu Leu His Asn Glu Ala Leu Ile
1 5

<210> SEQ ID NO 42
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 205-213 of UGT1A5 as
shown in Figure 4

<400> SEQUENCE: 42

Leu Leu His Asn Glu Ala Leu Ile
1 5

<210> SEQ ID NO 43
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 7-15 of UGT1A6 as
shown in Figure 4

<400> SEQUENCE: 43

Arg Ser Phe Gln Arg Ile Ser Ala Gly
1 5

<210> SEQ ID NO 44
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: amino acid residues 127-134 of UGT1A6 as
shown in Figure 4

<400> SEQUENCE: 44

Leu Leu Gln Asp Arg Asp Thr Leu
1 5

<210> SEQ ID NO 45
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 205-213 of UGT1A6 as
shown in Figure 4

<400> SEQUENCE: 45

Gln Arg Val Ala Asn Phe Leu Val Asn
1 5

<210> SEQ ID NO 46
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE

-continued

<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 7-15 of UGT1A7 as
shown in Figure 4

<400> SEQUENCE: 46

Thr Gly Leu Leu Pro Leu Tyr Val Cys
1 5

<210> SEQ ID NO 47
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: amino acid residues 127-134 of UGT1A7 as
shown in Figure 4

<400> SEQUENCE: 47

Leu Phe Asn Asp Arg Lys Leu Val
1 5

<210> SEQ ID NO 48
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 205-213 of UGT1A7 as
shown in Figure 4

<400> SEQUENCE: 48

Glu Arg Val Trp Asn His Ile Met His
1 5

<210> SEQ ID NO 49
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 7-15 of UGT1A8 as
shown in Figure 4

<400> SEQUENCE: 49

Thr Ser Pro Ile Pro Leu Cys Val Ser
1 5

<210> SEQ ID NO 50
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: amino acid residues 127-134 of UGT1A8 as
shown in Figure 4

<400> SEQUENCE: 50

Leu Phe Asn Asp Arg Lys Leu Val
1 5

<210> SEQ ID NO 51
<211> LENGTH: 9

-continued

<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 205-213 of UGT1A8 as
shown in Figure 4

<400> SEQUENCE: 51

Glu Arg Val Arg Asn His Ile Met His
1 5

<210> SEQ ID NO 52
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 7-15 of UGT1A9 as
shown in Figure 4

<400> SEQUENCE: 52

Thr Ser Pro Leu Pro Leu Cys Val Cys
1 5

<210> SEQ ID NO 53
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: amino acid residues 127-134 of UGT1A9 as
shown in Figure 4

<400> SEQUENCE: 53

Leu Phe Lys Asp Lys Lys Leu Val
1 5

<210> SEQ ID NO 54
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 205-213 of UGT1A9 as
shown in Figure 4

<400> SEQUENCE: 54

Glu Arg Val Arg Asn His Ile Met His
1 5

<210> SEQ ID NO 55
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 7-15 of UGT1A10 as
shown in Figure 4

<400> SEQUENCE: 55

Asp Gln Pro Arg Ser Phe Met Cys Val
1 5

-continued

```

<210> SEQ ID NO 56
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: amino acid residues 127-134 of UGT1A10 as
                          shown in Figure 4

```

```

<400> SEQUENCE: 56

```

```

Leu Phe Asn Asp Arg Lys Leu Val
1           5

```

```

<210> SEQ ID NO 57
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(9)
<223> OTHER INFORMATION: amino acid residues 205-213 of UGT1A10 as
                          shown in Figure 4

```

```

<400> SEQUENCE: 57

```

```

Glu Arg Val Trp Asn His Ile Val His
1           5

```

1. A method for predicting the potential risk and/or diagnosis of carcinomas or inflammatory bowel diseases on the basis of genetic disposition, characterized in that a DNA sample from a person to be investigated is tested for the presence of polymorphic UGT1A7 alleles which comprise mutations at codons 11, 129, 131 and/or 208.

2. The method as claimed in claim 1, where a positive result for a mutation at codon 11 and 208 is regarded as a positive indicator of a sensitivity for carcinomas, especially for colonic, pancreatic, hepatic, gastric and esophageal cancer.

3. The method as claimed in claim 2, where it is checked that the mutations regarded as positive are a W208R exchange and a silent mutation of CCC to CCA of codon 11.

4. The method as claimed in claim 1, where a positive result for a mutation at codon 11, 129, 131 and 208 is regarded as indicator of a sensitivity for an inflammatory bowel disease and a carcinoma.

5. The method as claimed in claim 4, where it is checked that the mutations regarded as positive are N129K, R131K and !208R [sic] and the silent mutation of CCC to CCA at codon 11.

6. The method as claimed in any of claims 1 to 5, characterized in that genomic DNA, preferably from lymphocytes, is used for the sample.

7. The method as claimed in claim 6, characterized in that the genomic deoxyribonucleic acid is isolated from a blood sample by methods of column chromatography and chemical processing.

8. The method as claimed in any of claims 1 to 7, characterized in that the DNA sample is used to carry out a PCR amplification of exon 1 of UGT1A7 (about 855 base pairs) with subsequent sequence analysis, and the sequence

found is compared with that of the wild type and of the polymorphic alleles of UGT1A7.

9. The method as claimed in any of claims 1 to 7, characterized in that cDNA fragments which match corresponding fragments of the wild-type allele or of the polymorphic alleles are generated in a polymerase chain reaction to the DNA sample with the aid of specific primer pairs, of which in each case one binds upstream and one binds downstream of the relevant mutated DNA regions around codons 11, 129/131 or 208, and in that the presence of mutations at codons 11, 129, 131 and/or 208 is detected with the aid of sequencing and/or hybridization techniques.

10. The method as claimed in claim 9, characterized in that the primer pairs bind in each case approximately 50 base pairs upstream and downstream of codon 11, of codon 129/131 and of codon 208, resulting in cDNA fragments about 100 to 150 base pairs in size.

11. The method as claimed in any of claims 8 to 10, characterized in that the presence of mutations at the relevant codons is detected by single strand conformation polymorphism analysis (SSCP) with a cDNA fragment, corresponding in each case, of the wild-type allele, where wild-type and polymorphic alleles are preferably distinguished by gel electrophoresis, and further preferably by temperature gradient gel electrophoresis (TGGE).

12. The method as claimed in claim 11, characterized in that the presence or absence of mutations is checked by sequencing the relevant cDNA by means of automated fluorescent dye sequencing.

13. The method as claimed in any of claims 8 to 10, characterized in that the presence of a polymorphism is

detected using a polymerase chain reaction with the aid of primer sequences which bind at their 3' terminus to the base change of the polymorphism.

14. The method as claimed in claim 13, characterized in that the presence of mutations is detected with the aid of homologous oligonucleotides having the wild-type or the polymorphism sequence and corresponding reverse primers (antisense primers) using a PCR technique.

15. The method as claimed in claim 14, characterized in that a quantitative Taqman PCR is used.

16. The method as claimed in claim 14 or **15**, characterized in that the PCR conjugates are detected using fluorescent dyes or enzyme markers.

17. A test arrangement comprising the genetic detection reagents necessary in a method as claimed in any of claims 1 to 15, namely the necessary primers or cDNAs, on a stationary support in an arrangement or sequence prepared for reading the result, where the DNA sample(s) prepared and additionally labeled according to in [sic] any of claims 8 to 10, of an individual to be investigated binds/bind to one of the detection reagents on bringing into contact with the test arrangement, and is/are detected with the aid of the marker at this site.

18. The test arrangement as claimed in claim 17, characterized in that the labeling is effected by fluorescent dyes or radioactive isotopes.

19. The test arrangement as claimed in claim 17 or **18**, characterized in that the detection reagents for a combined analysis of a plurality of UGT1A7 polymorphisms are immobilized together on a test arrangement.

20. A test arrangement as claimed in any of claims 17 to 19, characterized in that an inscription or an imprint assigned to the detection reagents is arranged on the stationary support for reading the result.

21. The use of polymorphic UGT1A7*2, UGT1A7*3 and UGT1A7*4 genes for preparing the polymorphic UGT isoforms encoded by them.

22. The use of the UGT isoforms as set forth in claim 21 for the metabolic characterization of tumor therapeutic agents and for investigating the toxicity and/or carcinogenicity of potential UGT1A7 substrates.

23. The use of UGT1A wild-type genes or gene fragments for producing a medicament for gene therapy treatment in cases of UGT1A polymorphisms, especially those of the UGT1A7*2 or UGT1A7*4 type.

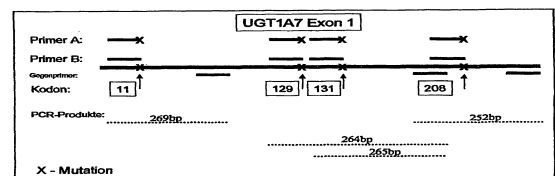
24. The use of recombinant UGT1A7 enzymes for therapeutic purposes.

* * * * *

专利名称(译)	预测癌症和炎性肠病的潜在风险的方法和相应的试验		
公开(公告)号	US20040121327A1	公开(公告)日	2004-06-24
申请号	US10/250508	申请日	2002-01-03
[标]申请(专利权)人(译)	曼斯MICHAEL 斯特拉斯堡CHRISTIAN		
申请(专利权)人(译)	曼斯MICHAEL 斯特拉斯堡CHRISTIAN		
当前申请(专利权)人(译)	MEDIZINISCHE Hochschule的HANNOVER		
[标]发明人	MANNS MICHAEL STRASSBURG CHRISTIAN		
发明人	MANNS, MICHAEL STRASSBURG, CHRISTIAN		
IPC分类号	G01N33/53 A61K38/45 A61K48/00 A61P3/04 A61P35/00 A61P39/02 A61P43/00 C07K14/47 C12M1/00 C12M1/34 C12N9/10 C12N15/09 C12Q1/48 C12Q1/68 C12Q1/6886 G01N27/447 G01N33/566 G01N33/58 G01N33/60 C07H21/04		
CPC分类号	C12Q1/6886 C07K14/47 C12Q2600/156 C12Q2600/172		
优先权	10100238 2001-01-05 DE		
其他公开文献	US7264928		
外部链接	Espacenet USPTO		

摘要(译)

本发明涉及预测癌症疾病或炎性肠病的潜在风险和/或诊断的方法，其中测试DNA样品中多态性UGT1A7等位基因的存在。突变的阳性结果是对癌症疾病敏感性的阳性指示。可以类似地预测对炎性肠病的敏感性。借助于DNA样品进行外显子1的PCR扩增，随后进行序列分析，并在该方法中进行，并将测定的序列与野生型和多态性等位基因进行比较。通过使用自动荧光染料测序对相应cDNA进行测序来监测突变的存在或缺失。所述方法的测试装置需要在预先制备的装置或序列中的固定支持物上的遗传检测试剂，即所需的引物或cDNA，以便读出结果。重组UGT1A7酶也用于治疗目的。



Figur 1

Allele		NC	CRC	IBD
UGT1A7*1	wildtype	48%	4%	6%
UGT1A7*2	+/+	0%	0%	0%
UGT1A7*2	+/-	5%	19%*	3%
UGT1A7*3	+/+	16%	0%	10%
UGT1A7*3	+/-	18%	23%*	10%
UGT1A7*4	+/+	7%	23%*	26%*
UGT1A7*4	+/-	3%	19%*	16%*
UGT1A7*3/ UGT1A7*4	compound heterozygous	3%	12%	16%*
UGT1A7*4 (+/+) and UGT1A7*3/UGT1A7*4		10%	35%*	50%*

Figur 2