Human Thiopurine Methyltransferase Pharmacogenetics

Kindred with a Terminal Exon Splice Junction Mutation That Results in Loss of Activity

Diane M. Otterness, Carol L. Szumlanski, Thomas C. Wood, and Richard M. Weinshilboum
Department of Pharmacology, Mayo Medical School/Mayo Clinic/Mayo Foundation, Rochester, MN 55905

Abstract

Thiopurine methyltransferase (TPMT) catalyzes *S*-methylation of thiopurine drugs such as 6-mercaptopurine. Large variations in levels of TPMT activity in human tissue can result from a common genetic polymorphism with a series of alleles for low activity. This polymorphism is an important factor responsible for large individual variations in thiopurine toxicity and therapeutic efficacy. We now report a new variant allele, *TPMT*4*, that contains a G→A transition that disrupts the intron/exon acceptor splice junction at the final 3′ nucleotide of intron 9, the terminal intron of the TPMT gene. This new allele cosegregated within an extended kindred with reduced TPMT activity. We attempted to determine the mechanism(s) by which the presence of *TPMT*4* might result in low enzyme activity. Although very few mature transcripts derived from allele *TPMT*4* were detected, the mutation did lead to generation of at least two aberrant mRNA species. The first resulted from use of a novel splice site located one nucleotide 3′ downstream from the original splice junction. That mRNA species contained a single nucleotide deletion and a frameshift within exon 10, the terminal exon of the gene. The second novel mRNA species resulted from activation of a cryptic splice site located within intron 9, leading to inclusion of 330 nucleotides of intron sequence. That sequence contained a premature translation termination codon. *TPMT*4* is the first reported allele for low TPMT activity as a result of a mutation within an intron. These observations also provide insight into mechanisms of mRNA processing after disruption of a terminal exon splice junction. (*J. Clin. Invest.* 1998. 101:1036–1044.) Key words: methyltransferase • pharmacogenetics • thiopurines • genetic polymorphism • RNA processing

Introduction

Thiopurine methyltransferase (TPMT; EC 2.1.1.67)[1] is a cytosolic enzyme that catalyzes the *S*-methylation of aromatic and heterocyclic sulfhydryl compounds, including the thiopurine drugs 6-mercaptopurine (6-MP) and 6-thioguanine (1–3). The immunosuppressant agent azathioprine is converted to 6-MP in vivo (4), so the active metabolite of azathioprine is also a substrate for TPMT. Thiopurines are used to treat patients with neoplasia and autoimmune disease as well as recipients of transplanted organs (4). Levels of TPMT activity and immunoreactive protein in human tissue are controlled by a common genetic polymorphism that is an important factor responsible for large individual differences in thiopurine toxicity and therapeutic efficacy (5–9). Allele frequencies for this polymorphism are such that ~ 1 in 300 Caucasians is homozygous for the allele or alleles for the trait of very low activity, ~ 11% of subjects are heterozygous and have intermediate activity, and ~ 89% are homozygous for the allele or alleles for high enzyme activity (5). Patients with genetically very low or undetectable levels of TPMT activity develop severe myelosuppression when treated with standard doses of thiopurines, while the same doses may undertreat patients with very high levels of enzyme activity (10–16). Because of its important clinical consequences, determination of TPMT phenotype in the red blood cell (RBC) has become a standard clinical pharmacogenetic test in some centers (17). Levels of TPMT activity in the RBC are highly correlated with relative levels of the enzyme activity in other human tissues and cells such as the kidney, liver, and lymphocyte (2, 18, 19).

TPMT cDNAs have been cloned from T84 human colon carcinoma cells and from human liver (20, 21). A TPMT-processed pseudogene with a sequence that is 96% identical to that of the open reading frame (ORF) of the TPMT cDNA has also been cloned and mapped to human chromosome band 18q21.1 (21). The active gene for the enzyme is ~ 34 kb in length, consists of 10 exons, and has been localized to chromosome band 6p22.3 (22). The wild-type allele for high TPMT activity has been designated *TPMT*1* (22), and to date eight variant alleles for very low TPMT activity, including the allele described here, have been reported (22–25). The most common of these in Caucasians, *TPMT*3A*, represents 55–70% of all variant alleles for very low TPMT activity (24, 25). *TPMT*3A* contains two point mutations, G460→A and A719→G, resulting in Ala154→Thr and Tyr240→Cys amino acid substitutions, respectively (see Fig. 1; 22, 24). When cDNAs containing either or both of these mutations were transiently expressed in COS-1 cells, the presence of each independently as well as both together resulted in striking decreases in the expression of both TPMT enzymatic activity and immunoreactive protein (22). Each of these two polymorphisms, that involving nucleotide 460 and that involving nucleotide 719, can also occur independently (25). The allele that contains only the G460→A polymorphism has been designated *TPMT*3B*, while that with only the A719→G polymorphism has been designated *TPMT*3C* (see Fig. 1; 25, 26). In the course of studies of the molecular basis for the TPMT genetic polymorphism, we identified a subject with very low activity who was heterozygous for *TPMT*3A,* and a new variant allele that contained a point mutation located at the final intron/exon 3′ splice junction of the

Address correspondence to Dr. Weinshilboum, Department of Pharmacology, Mayo Medical School/Mayo Clinic/Mayo Foundation, Rochester, MN 55905. Phone: 507-284-2246; FAX: 507-284-9111.

Received for publication 23 June 1997 and accepted in revised form 12 December 1997.

1. *Abbreviations used in this paper:* 6-MP, 6-mercaptopurine; ORF, open reading frame; RBC, red blood cell; RT, reverse transcription; TPMT, thiopurine methyltransferase.

J. Clin. Invest.

0021-9738/98/03/1036/09 $2.00
Volume 101, Number 5, March 1998, 1036–1044
http://www.jci.org

gene. In this study we set out to determine whether this new variant allele cosegregated with low TPMT activity within an extended pedigree, and, if so, to study the mechanism by which the mutation resulted in decreased enzyme activity.

Methods

Blood sample acquisition and preparation. Blood samples were obtained from ten members of an extended pedigree. Phenotypes for the TPMT genetic polymorphism had been determined previously in these subjects by measuring TPMT activity in RBC lysates, and are listed in the pedigree shown in Fig. 2 (5). All of the subjects were of Northern European ancestry. Genomic DNA was isolated from whole blood samples by use of the QIAamp Blood Kit (QIAGEN, Inc., Chatsworth, CA). This study was approved by the Mayo Clinic Institutional Review Board, and informed consent was obtained from all participants.

TPMT enzymatic assay. TPMT activity was measured with the radiochemical enzymatic assay of Weinshilboum et al. (27), an assay that is based on methylation of 6-MP with [^{14}C-methyl]-*S*-adenosyl-L-methionine as the methyl donor. One unit of enzyme activity represented formation of 1 nmole of 6-methylmercaptopurine per hour of incubation.

TPMT polymorphism detection and characterization. Individual exons of the TPMT gene were amplified by use of the PCR with intron-based primers to avoid amplification of the TPMT processed pseudogene (21). PCR amplification products were sequenced using dye primer cycle-sequencing chemistry to facilitate detection of heterozygous samples (28). PCR amplifications and DNA sequencing were performed exactly as described by Otterness et al. (25). For purposes of this report, the TPMT gene sequence of samples with high levels of enzyme activity reported by Szumlanski et al. (22) was defined as the wild-type.

TPMT intron 9 sequence. TPMT intron 9 was sequenced from a pBluescript SK subclone of the Lawrist 4 cosmid human genomic DNA clone that had been used to determine the TPMT gene sequence (22). The subclone was sequenced directly with the ABI PRISM Dye Terminator Cycle Sequencing Kit (Applied Biosystems, Inc., Foster City, CA). *TPMT* GenBank accession no.s U30517 and U30518 have now been updated to include the 1352-bp sequence of intron 9. Intron 9 was also sequenced by dye primer chemistry using DNA isolated from the proband in which the allele *TPMT*4* was discovered.

Reverse transcription(RT)-PCR and RFLP analysis. 8 ml of whole blood was obtained from the index subject, from a homozygous wild-type control subject, and from a control subject heterozygous for the nucleotide 719 polymorphism (i.e., a subject with *TPMT* genotype **1/*3C*; see Fig. 1). Lymphocytes were isolated from each of these blood samples by Ficoll-Paque density gradient centrifugation (Pharmacia LKB Biotechnology Inc., Piscataway, NJ), and total RNA was isolated from the lymphocytes with the Qiagen RNeasy kit (QIAGEN Inc.). The RNA samples were treated with DNase I (Boehringer Mannheim Biochemicals, Indianapolis, IN), and first-strand cDNA was synthesized with the Pharmacia First Strand cDNA Synthesis kit (Pharmacia LKB Biotechnology Inc.). The primers used to synthesize cDNA were either a *TPMT*-specific reverse primer, R971, a primer that anneals to the 3′ untranslated region of the TPMT mRNA, or the Not I-d(T)$_{18}$ primer provided with the kit (Table I). The numbering scheme for all primers, with the exception of those for introns, was based on the TPMT cDNA sequence. The A in the ATG translation initiation codon was at position +1. The cDNA synthesized with these primers was then used as template for the PCR. Initially, primers F515 and R971, primers that anneal to sequence within *TPMT* exons 8 and 10, respectively, were used to perform RT-PCR with first-strand cDNA generated with primer R971. Amplification products were subcloned into the vector pCR2.1 (Invitrogen Corp., San Diego, CA). DNA was isolated from overnight cultures with the Wizard miniprep DNA isolation kit (Promega Corp., Madison, WI) and was digested with 5 U of Acc I (Boehringer Mannheim Biochemicals) for 4 h at 37°C. Acc I-resistant subclones were then sequenced.

A similar RT-PCR amplification was performed with the same primers, but in this case the first-strand cDNA was generated with the primer Not I-d(T)$_{18}$. Amplification products from these reactions were isolated by agarose gel electrophoresis, extracted from the gel with the Gene Clean kit (BIO 101, La Jolla, CA), and subjected to Acc I (5 U) digestion for 4 h at 37°C. The DNA fragments were separated by electrophoresis on a 3.5% NuSieve Gel (FMC Bioproducts, Rockland, ME). Undigested product amplified from the index subject's cDNA was extracted from the gel, and was subcloned into pCR2.1 before sequencing.

In a separate series of experiments, TPMT primers F231 and R523, primers that anneal to sequence within *TPMT* exons 5 and 8, respectively, were used to perform RT-PCR with first-strand cDNA generated with the Not I-d(T)$_{18}$ primer from the index subject, and from a homozygous wild-type control subject. A *TPMT*3A* cDNA expression construct was also used as template for the PCR to generate a control amplification product for Mwo I restriction digestion. These amplification products were isolated by agarose gel electrophoresis, extracted with the Gene Clean kit, and subjected to digestion for 4 h at 60°C with 5 U of Mwo I (New England Biolabs Inc., Beverly, MA). The DNA was then analyzed on a 3.5% NuSieve Gel.

TPMT primers F231 and I9R977 or I9R1305, primers that anneal to sequence within exon 5 and intron 9, respectively, as well as I9F1298 and NotI-d(T)$_{18}$, primers that anneal to *TPMT* intron 9 sequence and to the cDNA poly (A) tract, respectively, were also used to perform RT-PCR. Products obtained after amplification performed with primer pair F231, and I9R1305 or primer pair I9F1298 and Not I-d(T)$_{18}$ were subcloned into pCR2.1 without prior agarose gel purification, and positive colonies were analyzed to determine which might contain appropriate inserts by performing the PCR with the primers used originally to generate the subcloned inserts. DNA was then isolated from overnight cultures of positive colonies, and was sequenced.

Finally, Southern blot analysis (29) was performed with RT-PCR products that had been generated with primer pair F515 and R971 or with primer pair F515 and I9R1305 to allow a semiquantitative estimate of the concentration of *TPMT*4* mRNA as compared with that derived from allele **3A*. To perform these experiments, primer R625 was 5′ end-labeled with [γ-^{32}P]ATP (ICN Biomedicals, Inc., Irvine, CA) by use of T4 polynucleotide kinase (Promega Corp.). Hybridization was performed with ExpressHyb (Clontech, Palo Alto, CA) according to the manufacturer's protocol. Serial dilutions of the RT-PCR product generated with primer pair F515 and R971 served as a standard curve that was used to estimate relative quantities of **3A* and **4* RT-PCR amplification products. The Southern blots were analyzed by densitometry performed with the Ambis optical system (Scananalytics, Billerica, MA). Exposure times for x-ray film were then adjusted to obtain a linear standard curve within the concentration range of interest.

Results

Segregation analysis

The original report of inheritance of levels of RBC TPMT activity included two kindreds and one family identified on the basis of a proband homozygous for the trait of very low RBC TPMT activity (5). We returned to those pedigrees to determine the genotypes of proband subjects after DNA sequences for a series of variant alleles for very low TPMT activity had been identified (25, 26). The probands in families FA35 and FA37 were both homozygous for *TPMT*3A*, the most common variant allele in Caucasians (22, 25). However, the

Table I. Sequences of Primers Used to Perform PCR Amplifications

Primer designation	Primer location	Primer sequence
Exon sequence determination		
−21M13F	—	TGTAAAACGACGGCCAGT
M13R	—	CAGGAAACAGCTATGACC
I2F (−87)	Intron 2	ACTGCTAAGAATAATAGGTTTTCATTTAGTTC
I3R153	Intron 3	GCCACAGATGCACTGTGACTCGGGAG
I3F (−179)	Intron 3	TACCACTGACTGGGTGTGTGTCTGA
I4R207	Intron 4	CTCAATCCAGAAAGACTTCATACCTGTT
I4F (−141)	Intron 4	CCTGCATGTTCTTTGAAACCCTATGAA
I5R234	Intron 5	GCTTGAGTACAGAGAGGCTTTGACCTC
I5F (−224)	Intron 5	TGTCCTCTGTGATATTCCTCTGAGTTG
I6R110	Intron 6	GTGGATGTTACACAGGAGGAAGAGAG
I6F467	Intron 6	CTCCACACCCAGGTCCACACATT
I7R54	Intron 7	GTATAGTATACTAAAAAATTAAGACAGCTAAAC
17F (−199)	Intron 7	CCCAGCTTAGGCAGGGGCCATAA
I8R146	Intron 8	TCCAAACTGGAATTATCTCCATGTA
I8F (−108)	Intron 8	GAGAAGAACATGCCACATCATCACCTA
I9R38	Intron 9	TTTGTTTAAAAAGTTACAGCATAAGT
I9F1298	Intron 9	AATCCCTGATGTCATTCTTCATAGTATTT
R971	Exon 10	CACATCATAATCTCCTCTCC
Intron 9 sequence determination		
I8F (−108)	Intron 8	GAGAAGAACATGCCACATCATCACCTA
I9F96	Intron 9	TGGGAGGCCAAGGTGGGCAGATCA
I9F347	Intron 9	AACCTATTGTGAACTATTAG
I9F463	Intron 9	TATGCTCTGTATATATGCTATATG
I9F829	Intron 9	GCCCAGTGGTGTATTTATTTTTACTC
I9F1087	Intron 9	CCTGCCTCAACCTCCCAAGTAGCT
I9R247	Intron 9	CCATGCCTCAGCCTCTCAAATAGT
I9R503	Intron 9	ATACCTTTAACATAGAGCATATAG
I9R608	Intron 9	GAGCTGTGATTGCACTTCTGTACT
I9R830	Intron 9	GCATGGCGGCTCATGCCTCTAATC
I9R1189	Intron 9	GCCTGGCCAAGATGGTGAAACTCT
I9R1305	Intron 9	CAGGGATTCTTTTAAAAATACTC
R821	Exon 10	CATCCATTACATTTTCAGGCTTTAGCATAAT
RT-PCR		
F515	Exon 8	CCCTCCTGGGAAAGAAGTTTCAGTATC
R971	Exon 10	CACATCATAATCTCCTCTCC
F231*	Exon 5	ATGGTTTGCAGACCGGGGAC
R523*	Exon 8	CCAGGAGGGAAAACATTGTATCTG
R625*	Exon 9	CAAACAACCTTTCAATTTCAGC
I9R977	Intron 9	CCTTGAAATTATTTTAAAAACAG
I9R1305*	Intron 9	CAGGGATTCTTTTAAAAATACTC
I9F1298	Intron 9	AATCCCTGATGTCATTCTTCATAGTATTT
Not I-d$(T)_{18}$	Poly (A) tract	AACTGGAAGAATTCGCGGCCGCAGGAAT$_{18}$

Eighteen nucleotides of the −21M13F or M13R sequences were added to the 5′ ends of each forward and reverse primer, respectively, to make it possible to use dye primer DNA sequencing chemistry (see Methods). In the abbreviations used to designate primers, *I*, intron; *F*, forward; and *R*, reverse. *Indicates those primers that did not contain M13 sequences.

proband in kindred FA45 was heterozygous for the allele **3A* polymorphisms at nucleotides 460 and 719, and had no other nucleotide polymorphisms within the ORF. Therefore, either this subject carried a new variant allele that affected nucleotides outside of the ORF, or he was a compound heterozygote for *TPMT*3B/*3C*, i.e., he carried the nucleotide 460 and 719 polymorphisms on separate alleles (Fig. 1). Either of those situations would result in very low TPMT activity. In an attempt to determine the cause of very low enzyme activity in this subject, we sequenced all *TPMT* exons that encode protein (exons 3–10) as well as their intron/exon splice junctions using DNA from the proband and his immediate family members. DNA sequencing was performed with dye primer rather than dye terminator chemistry to facilitate detection of heterozygous sequence (28).

Use of this approach led to discovery of a new point mutation located at the intron 9/exon 10 splice junction, which consisted of a G→A transition at the final nucleotide of the intron.

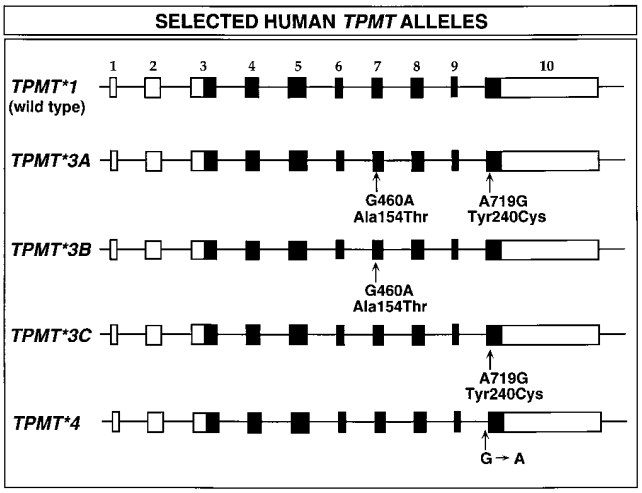

Figure 1. Selected human *TPMT* alleles. The figure depicts schematically the wild-type allele for *TPMT* (*TPMT*1*) as well as four variant alleles for very low enzymatic activity. *Black rectangles* represent exons that encode ORF sequence, while *white rectangles* represent exons or portions of exons that encode untranslated region sequence. *Numbers above the rectangles* represent exon numbers. Exon sizes are proportional to the relative lengths shown, but introns vary greatly in length (22) even though they have been depicted here as equal in length.

Furthermore, this mutation cosegregated within the kindred with reduced TPMT activity, and it segregated independently of the two **3A* polymorphisms at nucleotides 460 and 719 (Fig. 2), confirming its location on an alternative allele and demonstrating that the proband was not a compound heterozygote for alleles **3B/*3C*. We designated this new allele *TPMT*4*. Both the proband and his brother had very low RBC TPMT enzymatic activity, and both were compound heterozygotes for alleles **3A* and **4* (Fig. 2). All members of the third generation of this family had intermediate TPMT enzymatic activity ($<$ 9.5 U/ml RBC), and all of them were heterozygous for either allele **3A* or **4* paired with the wild-type allele (Fig. 2). Because subject I-1, the grandfather in the family, was heterozygous for alleles **1* and **3A*, we concluded that allele **4* had been inherited from subject I-2, the grandmother, a subject with very low RBC TPMT activity (Fig. 2). Unfortunately she was deceased, so we were unable to determine her genotype, although her level of RBC TPMT activity had been measured in the course of the original RBC TPMT population and family studies (5). Both *TPMT* genotypes and levels of RBC TPMT activity for members of this kindred are shown in Fig. 2. The next step in our analysis was to study the molecular consequences of the *TPMT*4* allele, i.e., the mechanism(s) by which it resulted in low TPMT enzymatic activity.

*TPMT*4 molecular mechanisms*

Exon skipping is the most common consequence of mutations that disrupt splice sites (30). However, the *TPMT*4* mutation affected the 3′-end of the final intron of the gene, making it impossible in the usual sense to skip the affected exon and proceed to the next. Therefore, we set out to study the functional consequences of this particular splice site mutation. The strategy involved use of RT-PCR performed with mRNA isolated from peripheral blood mononuclear cells; cells known to express TPMT, and in which the relative level of enzyme activity is correlated with that expressed in the RBC (18). However, since the index subject was a compound heterozygote for alleles **4* and **3A*, it was necessary to be able to discriminate mRNA species derived from those two alleles. To make that possible, we took advantage of the creation or disruption of restriction sites by the two polymorphisms present in allele **3A*. The two **3A* polymorphisms bracket the location of the **4* mutation, i.e., they are located on either the 3′ or 5′ side of the **4* mutation. Specifically, the allele **3A* nucleotide 719 polymorphism creates an Acc I restriction site that is not present in either the wild-type allele or in allele **4*, a restriction site that is located downstream of the **4* mutation. Conversely, the **3A* polymorphism at nucleotide 460 destroys a Mwo I restriction site that is present in both the wild-type allele and in allele **4*, a restriction site located upstream of the **4* mutation. Since exon skipping in the usual sense could not take place with the *TPMT*4* splice site mutation, alternative mechanisms had to be considered. Included among possible functional consequences of the **4* mutation would be use of a novel splice site created by the mutation itself, use of an alternative polyadenylation signal located within intron 9 (i.e., skipping of exon 10 by excluding it from the terminus of the mRNA), or activation of a cryptic splice site located within *TPMT* intron 9 that is not usually used. A series of RT-PCR experiments was then performed to search sequentially for one or a combination of these possible functional consequences of the *TPMT*4* mutation.

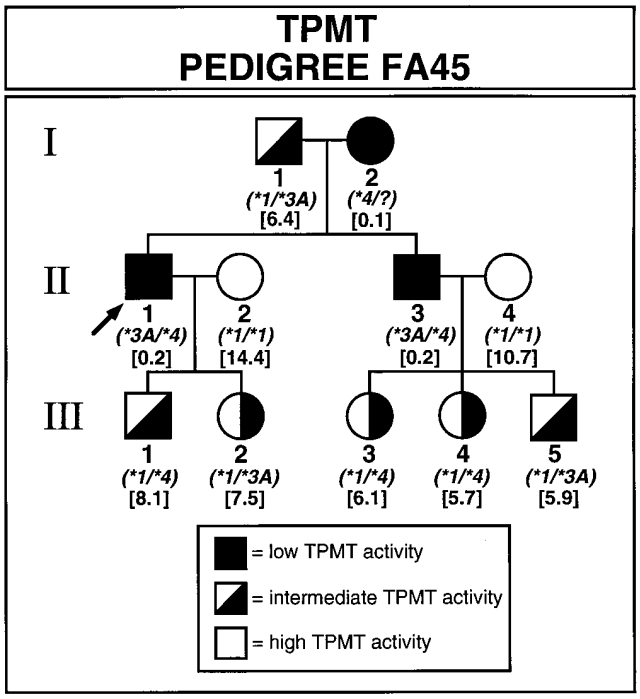

*Figure 2. TPMT*4*, **3A,* and **1* segregation within a pedigree. *TPMT* genotypes are indicated in *parentheses*, and RBC TPMT enzyme activities in U/ml packed RBCs are shown in *brackets*. *Circles* represent women, and *squares* represent men. The proband is indicated with an *arrow*.

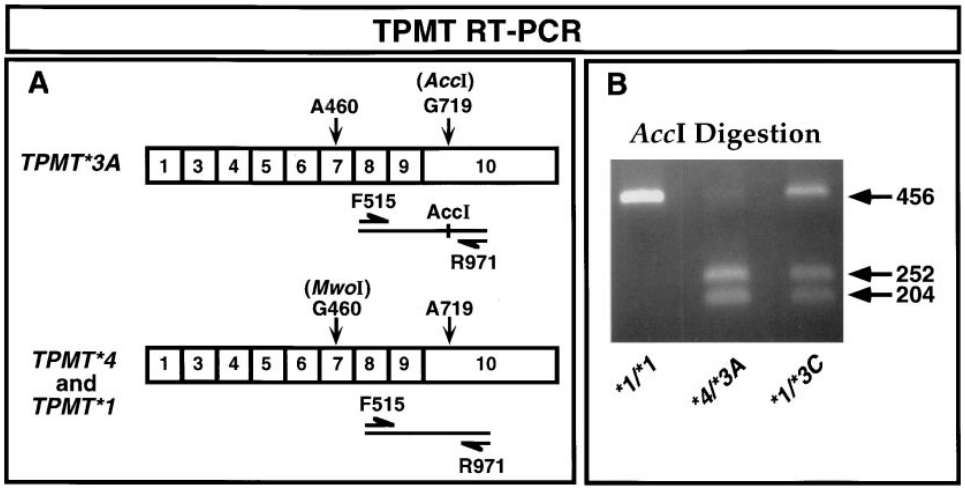

Figure 3. TPMT RT-PCR and RFLP analyses. (*A*) Schematic depiction of the RT-PCR and Acc I RFLP analyses of *TPMT*4*, **3A,* and **1*. In this and subsequent figures, *numbers within the rectangles* refer to exon numbers. Lengths of exons 1–9 are depicted as being equal, although they differ in length (see Fig. 1). Exon 2 is not included because, as a result of alternative splicing, that exon is represented infrequently in mature TPMT mRNA (22). (*B*) Acc I endonuclease digestion of RT-PCR products from a wild-type control subject (**1/*1*); the proband (**4/*3A*); and a heterozygous control subject (**1/*3C)*. See text for details.

Novel splice site usage. The point mutation present in *TPMT*4* created a novel splice site since the first nucleotide in exon 10 is a G. Therefore, the mutation created a new AG sequence located one nucleotide 3′ of the original AG (Table II). To determine whether splicing of mRNA occurred at this point in transcripts derived from *TPMT*4*, RT-PCR was performed with first-strand cDNA prepared from the proband, and with primers F515 and R971, primers that annealed to *TPMT* exons 8 and 10, respectively. Use of these primers made it possible to use Acc I digestion of the nucleotide 719 polymorphism within **3A* exon 10 to differentiate between mRNA derived from alleles **3A* and **4* (Fig. 3 *A*). The amplification products obtained from this reaction were subcloned and analyzed by restriction digestion performed with AccI. Analysis of 98 subclones revealed only one that contained the product expected from a **4* transcript, a product in which exon 10 was shortened by a single nucleotide as a result of use of the novel splice site created by the **4* mutation (Table II). Therefore, although use of the novel splice site did occur, it appeared to be a rare event. We tested that conclusion further by performing a second independent experiment.

The strategy used in the second experiment involved an attempt to enhance the capture of subclones derived from transcripts that used the novel splice site created by the **4* mutation. In this experiment, RT-PCR products generated with primer pair F515 and R971 were subjected to Acc I digestion before rather than after subcloning. In addition, a sample homozygous for the wild-type allele that lacks the Acc I restriction site, and a sample from a subject heterozygous at nucleotide 719 (genotype **1/*3C*; Fig. 1) and, therefore, heterozygous for the Acc I restriction site, were included as controls for the restriction digestion. In this case, the wild-type control sample, which lacks the Acc I site, would be expected to generate a DNA fragment 456 bp in length, while the samples heterozygous for either **1* or **4* with **3A* would be expected to have three fragments 456, 252, and 204 bp in length (Fig. 3 *B*). Restriction digestion of the RT-PCR amplification product generated with cDNA from the proband predominantly yielded two cleavage fragments 252 and 204 bp in length (Fig. 3 *B*). However, a faint uncleaved product remained after digestion with Acc I. This 456-bp product was isolated by agarose gel electrophoresis, and was subcloned to determine whether it was derived from **4*. Sequencing demonstrated that DNA from these subclones contained only products generated from allele **3A*. Therefore, the faint 456-bp band shown in the lane labeled **4/*3A* in Fig. 3 *B* apparently resulted from incomplete Acc I digestion of the RT-PCR product. These observations supported the results obtained during the initial experiment, and confirmed that use of the novel splice site generated by the **4* mutation was a rare event.

Table II. TPMT Intron 9/Exon 10 Acceptor Splice Sites

	3′ Splice site consensus sequence															
	−14	−13	−12	−11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	+1	Position
	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	C	A	G	G	Nucleotide
3′ Splice sites	78	81	83	89	85	82	81	86	91	87		78	100	100	55	Percentage
WT 1339*	C	T	T	T	C	T	T	G	T	T	T	C	A	G	G	
Novel 1340*	T	T	T	C	T	T	G	T	T	T	C	A	A	G	T	
Cryptic 1009*	C	T	C	T	T	G	T	T	G	C	C	T	A	G	G	

The consensus 3′ splice site sequence (38) is compared with that of the wild-type *TPMT*1* allele (*WT*) as well as the novel splice site created by the *TPMT*4* mutation and the cryptic splice site that can be activated within intron 9 in the presence of the *TPMT*4* mutation. *Y* designates a pyrimidine, and *N* designates any base. The invariant AG splice site sequence is *boxed*. *Dotted lines* connect identical nucleotides, locations of which have been displaced by the *TPMT*4* mutation. *Indicates the location of the initial intron 9 nucleotide for the sequence listed.

They also indicated that mature transcripts derived from allele **4*, at least transcripts that included sequence extending from exon 8 to exon 10, were rare. The next functional consequence of the **4* mutation tested was the possible use of a polyadenylation signal located within intron 9, resulting in exclusion of exon 10 from the mature mRNA.

Alternative polyadenylation site usage. Use of a polyadenylation signal located within intron 9 would result in termination of the transcript within intron 9 and, therefore, in exclusion of exon 10 from the transcript. The RT-PCR primers used to perform the preceding set of experiments, primers that annealed within exons 8 and 10, would fail to detect transcripts that terminated within intron 9. To determine whether such mRNA species might be generated, RT-PCR was performed with cDNA prepared from the proband as well as a control cDNA sample from a subject homozygous for the wild-type allele and an additional control for allele **3A* (see Methods). In this case, the primers used were F231 and R523, primers that anneal to exons 5 and 8, respectively. The template cDNA for these experiments had been generated with the NotI-d$(T)_{18}$ primer. Mwo I restriction digestion was used to distinguish between amplification products derived from alleles **1*, **3A*, and **4*. RT-PCR products generated from allele **4* would be cleaved by Mwo I. However, products amplified from allele **3A* would not be digested because the **3A* polymorphism at nucleotide 460 disrupts this restriction site (Fig. 4). As anticipated, the wild-type control sample was cleaved by Mwo I, and its digestion generated DNA fragments of the appropriate sizes. The **3A* control, which contained the nucleotide 460 polymorphism that disrupts the Mwo I site, was not cleaved. The product generated by amplification of DNA isolated from the proband also showed no evidence of restriction digestion with Mwo I (Fig. 4 *B*). Therefore, only RT-PCR products generated from allele **3A* could be detected in the sample from the proband. These results were compatible with the conclusion that activation of an alternative polyadenylation signal located upstream of the **4* splice site mutation is an infrequent event if it occurs at all. These observations were also compatible with the conclusion that very few mature **4* transcripts, at least transcripts that included *TPMT* exons 5–8, were present. The final possible functional consequence of the **4* mutation to be tested was activation of a cryptic splice site located within intron 9.

Cryptic splice site usage. Activation of a cryptic splice site is the second most frequent consequence of mutations that alter splice junctions (30). Since most cryptic splice sites are located relatively near the authentic splice site (30), we first needed to determine the entire DNA sequence of intron 9 to know whether it might contain potential 3′ splice site sequences. Because the original report of the TPMT gene structure (22) included only a portion of the sequence of intron 9, we sequenced the entire intron from both the *TPMT* genomic DNA clone used originally to determine the gene structure (22), and DNA isolated from the *TPMT*4* proband. Intron 9 consisted of 1352 nucleotides and included three *Alu* sequences. Other than the *TPMT*4* mutation at the 3′ splice junction, the only variant sequence within the intron of the proband was located at intron nucleotide 900. At that point, the proband's DNA was heterozygous for G/T, while the wild-type sequence was G. Segregation analysis confirmed that in this kindred, allele **4* had a T at nucleotide 900, while **3A* had a G at that position. Analysis of the DNA sequence of intron 9 revealed several possible cryptic 3′ splice sites and associated branch points, one of which was subsequently demonstrated to be capable of activation in allele **4*.

To test for cryptic splice site activation, RT-PCR was performed with cDNA from the proband as template and with primer F231 (exon 5–specific) paired with either intron 9–specific primer I9R977 or primer I9R1305 (reverse primers that began at intron 9 nucleotides 977 and 1305, respectively; Fig. 5). Neither of these intron 9–specific primers annealed to the *Alu* sequence. No detectable amplification product was obtained with primer pair F231 and I9R977. However, two amplification products ~ 490 and 670 bp in length, respectively, were obtained when primer F231 was paired with I9R1305 (Fig. 5). These two products were subcloned into pCR2.1, and 125 subclones were tested by performing the PCR with primer pair F231 and I9R1305 to determine whether they contained the expected inserts. Only six of the 125 clones were positive, and these six clones were subjected to restriction digestion with Eco RI to determine insert length. In all six cases, the insert was ~ 670 bp long. When these clones were sequenced, all

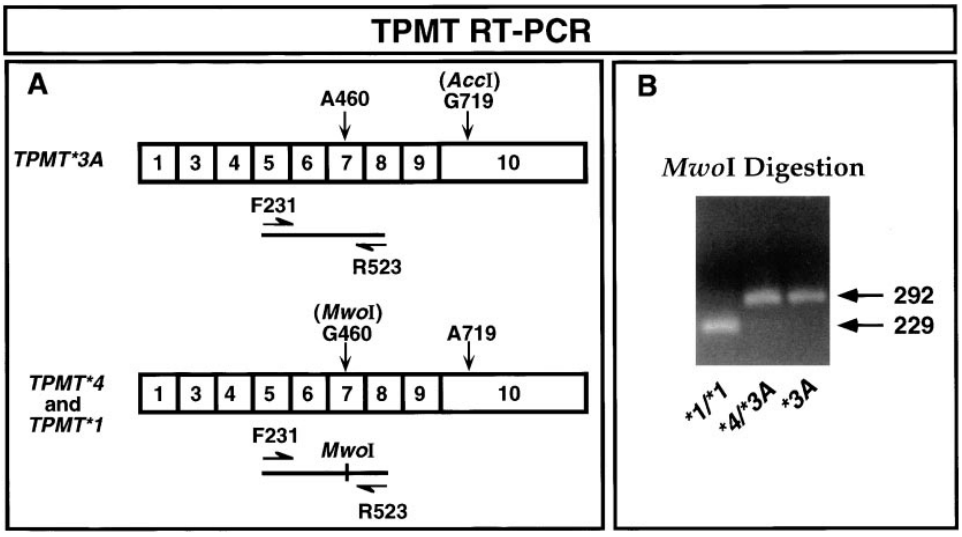

Figure 4. TPMT RT-PCR and RFLP analyses. (*A*) Schematic depiction of the of RT-PCR and Mwo I RFLP analyses of *TPMT*4*, **3A,* and **1*. (*B*) Mwo I endonuclease digestion of RT-PCR products from a wild-type control subject (**1/*1*); the proband (**4/*3A*); and a *TPMT*3A* sequence control (**3A*). See text for details.

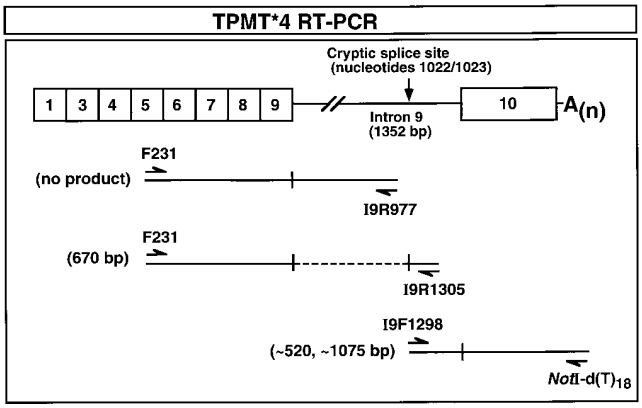

Figure 5. RT-PCR analyses performed with *TPMT* intron 9–specific primers. See text for details.

six contained *TPMT*4* sequence that included exons 5–9, plus a portion of intron 9 that extended from nucleotide 1023 to the location of the reverse primer (Fig. 5). Selected clones that had not tested positive during the analytical PCR performed with the original primers were also digested with Eco RI to determine insert size. A majority of those clones contained inserts that were ~ 490 bp in length, while a minority contained smaller inserts. DNA sequencing demonstrated that these subcloned products had been generated by nonspecific amplification of cDNA by primer F231 and residual Not I-d$(T)_{18}$ primer that had been carried over from the original first-strand cDNA synthesis reaction. None of these subclones showed sequence identity with *TPMT*. In summary, these experiments demonstrated activation of a cryptic 3′-acceptor splice site located within intron 9 of *TPMT*4*, with splicing between intron nucleotides 1022 and 1023. That location included one of the potential cryptic 3′ splice sites identified when we had sequenced intron 9 (see Table II for the DNA sequence of this region). This mRNA contained the 3′ end of *TPMT* intron 9 from nucleotide 1023 onward in the 3′ direction.

In an effort to confirm and extend these results, RT-PCR was also performed with an intron 9–specific forward primer, I9F1298 (i.e., a primer located within that portion of the intron retained as a result of cryptic splice site activation) paired with a Not I-d$(T)_{18}$ reverse primer (Fig. 5). This experiment was designed both to determine whether the usual *TPMT* polyadenylation site(s) was maintained when allele **4* underwent cryptic splice site activation, and to obtain the 3′ terminus of the resultant transcript. Performing this experiment would also confirm retention of intron 9 sequence within this mRNA species from nucleotide 1023 to the end of the intron. Two PCR products ~ 520 and 1075 bp in length (Fig. 5) were amplified with this set of primers when the template was cDNA produced from RNA isolated from the proband. We subcloned and sequenced these amplification products and confirmed that they contained a portion of intron 9 from nucleotides 1298–1352 (i.e., from the forward primer to the end of the intron), together with exon 10. These products had been derived from *TPMT*4* because the sequence at nucleotide 719 was that of **4* and not **3A*, and the original intron 9 nucleotide 1352 was A, not G (i.e., it was the mutant **4* splice site nucleotide). The presence of PCR products of two different lengths resulted from the use of two different polyadenylation signals located within exon 10, and corresponded to the shorter two of the three TPMT transcripts that have been observed for the wild-type allele in most human tissues (21). No subclones that contained transcripts generated from *TPMT*3A* were identified when intron 9-specific primers were used to perform either set of RT-PCR experiments, i.e., activation of a cryptic splice site occurred within intron 9 for allele **4*, but not allele **3A*. Finally, in an attempt to quantitate the *TPMT*4* transcript produced by cryptic splice site activation, we performed Southern blot analyses using RT-PCR products generated with either primer pair F515 and R971 or primer pair F515 and I9R1305. Primers F515 and R971 annealed to exons 8 and 10, respectively, and would be expected to result in an RT-PCR product 457 bp in length from normally spliced TPMT RNA. Primers F515 and I9R1305 annealed to exon 8 and within intron 9, respectively, and would be expected to result in an RT-PCR product 394 bp in length if the cryptic 3′ splice site within intron 9 were activated. The amplifications performed with primer pair F515 and R971 resulted in an RT-PCR product of the expected size from both the proband (genotype **3A/*4*) and from the wild-type control sample. Primer pair F515 and I9R1305 resulted in an RT-PCR product 394 bp in length only with cDNA prepared from the proband. The quantity of this cryptic splice site product was much less than that obtained with the other primer pair. As shown in Fig. 6, serial twofold dilutions (2–0.02 μl) of the product obtained with primer pair F515 and R971 were used to estimate the quantity of product obtained with primer pair F515 and I9R1305. Ten times the amount of the cryptic splice site product, in comparison to the highest concentration of the standard curve, was analyzed on the gel. These results demonstrated that mRNA, which contained the 3′ portion of intron 9, accounted for only ~ 4–7% of all TPMT transcripts (Fig. 6). No mRNA that contained intron 9 sequence was detected in a wild-type control sample (data

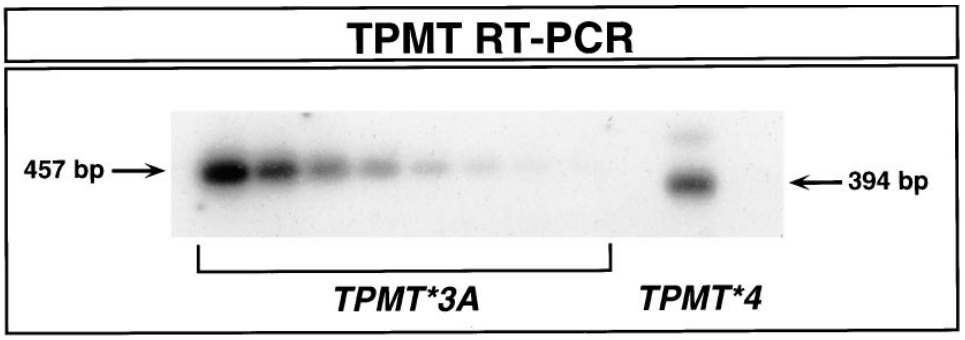

Figure 6. Southern analysis of TPMT RT-PCR products. Serial twofold dilutions (2–0.02 μl) of RT-PCR product obtained with the primer pair F515 and R971 for the proband subject with genotype *TPMT*3A/*4* are shown on the left. 20 μl (10× the highest dilution shown on the left) of RT-PCR product obtained with primer pair F515 and I9R1305 for the same subject is shown on the right. See text for details.

not shown). Translation of the **4* mRNA species that contained the 3′ portion of intron 9 would result in generation of a truncated protein because of the presence of a translation termination codon within intron 9 located one codon downstream of the end of exon 9. Therefore, if this rare mRNA were translated, the result would be deletion of the terminal 36 amino acids of the encoded protein as compared with the wild-type TPMT protein.

Discussion

The TPMT genetic polymorphism represents a striking example of the potential clinical importance of pharmacogenetic variation in expression of a drug-metabolizing enzyme (7–9). Individuals with genetically very low levels of TPMT activity are at a greatly increased risk for potentially life-threatening toxicity when exposed to standard doses of thiopurines, while those with very high levels of this enzyme activity may be undertreated with the same dosages of these drugs (10–16). A total of eight different variant alleles for low TPMT activity, including the **4* allele described here, have been reported to date (22–25). The most common variant allele for very low TPMT activity in Caucasians, *TPMT*3A*, has two point mutations, both of which alter amino acid sequence (22, 24). All other known variant alleles, with the exception of **4*, involve mutations within the ORF of TPMT mRNA (23, 25).

We have now characterized the first variant allele for very low TPMT activity, *TPMT*4*, that interrupts a splice junction sequence. This allele segregated within an extended kindred with low TPMT activity (Fig. 2). We also studied possible mechanisms by which *TPMT*4* might result in very low enzyme activity. This allele contains a G→A point mutation at the intron 9/exon 10 splice junction that disrupts the nearly invariant AG 3′ splice site sequence for eukaryotic organisms (31). Splice site mutations account for ~ 10% of all mutations listed in the Human Genome Mutation Database (32). However, mutations at 3′ acceptor splice sites are less frequent than are those at 5′ donor splice sites, and only a few 3′ splice site mutations involving a terminal exon have been studied mechanistically (30, 32). In the case of *TPMT*4*, we have demonstrated that disruption of the 3′ splice site sequence resulted in either use of a novel splice site located one nucleotide downstream of the original splice site, or activation of a cryptic 3′ splice site within the preceding intron, resulting in inclusion of 330 nucleotides of intron sequence within the mature *TPMT*4* transcript. Activation of a cryptic splice site is the second most frequent consequence (after exon skipping) of this type of mutation (30). Obviously, when the 3′ acceptor splice site of a terminal exon is involved with a mutation, exon skipping, at least in the usual sense, cannot occur. The distance between the original splice site and a cryptic site that is activated is usually ~ 100 bp (30). However, cryptic splice sites have been shown to occur up to 300–400 bp distant from the authentic site (30), as we found for *TPMT*4*. Translation of either of the *TPMT*4* transcripts that we identified would result in an aberrant TPMT protein. The transcript with the single nucleotide deletion as a result of creation of a novel splice site would contain a single nucleotide frameshift (Table II), and the final 38 amino acids of the encoded protein would differ from those of the wild-type protein. TPMT mRNA species that included a portion of intron 9 would contain a premature translation termination codon, and would encode a truncated protein. However, it should be emphasized that both of these *TPMT*4* transcripts were rare. For example, the mRNA species resulting from use of the novel splice site (Table II) was present in only ~ 1% of the clones studied, and the mRNA species resulting from activation of the cryptic splice site within intron 9 accounted for only ~ 4–7% of all transcripts. One possible explanation for their very low abundance might be instability of the aberrant mRNA species produced from **4*. Aberrant transcripts that contain premature termination codons have been shown to have decreased stability (33, 34). Obviously, we also cannot rule out the possibility that the intron 9/exon 10 mutation described here might be linked to another mutation located within promoter or enhancer elements that would result in decreased transcription. Very low levels of aberrantly spliced transcripts have also been shown to occur naturally for some wild-type genes (35–37). However, no *TPMT*3A* transcripts were detected when intron 9–specific primers were used to perform RT-PCR, suggesting that the cryptic splice site activated in allele **4* is not used during splicing of *TPMT*3A* RNA.

There are relatively few carefully studied examples of aberrant splicing in the presence of mutations which, like that in allele *TPMT*4*, are located at terminal intron/exon junctions. These observations also contribute to understanding mechanisms of splicing at the 3′-terminal exons of genes. *TPMT*4* has only been observed in this extended family. Therefore, the frequency of this allele in the population at large, if it is present outside of this kindred, has not been determined. However, because of the clinical significance of inherited variation in levels of TPMT activity, characterization of as many variant alleles responsible for very low TPMT activity as possible will be required so that DNA-based diagnostic tests can be compared with the phenotypic test presently used to individualize therapy with thiopurine drugs. The ultimate goal is to minimize toxicity and maximize the therapeutic efficacy of this important group of therapeutic agents.

Acknowledgments

We thank Luanne Wussow for her assistance with the preparation of this manuscript, and Dr. Eric Wieben for his helpful suggestions. We would also like to thank the members of family FA45 for their participation in this study.

Supported in part by National Institutes of Health grants RO1 GM28157 (R.M. Weinshilboum) and RO1 GM35720 (R.M. Weinshilboum), as well as a supplement to RO1 GM35720 supported by the Office of Research on Women's Health (R.M. Weinshilboum).

References

1. Remy, C.N. 1963. Metabolism of thiopyrimidines and thiopurines: S-methylation with S-adenosylmethionine transmethylase and catabolism in mammalian tissue. *J. Biol. Chem.* 238:1078–1084.

2. Woodson, L.C., and R.M. Weinshilboum. 1983. Human kidney thiopurine methyltransferase: purification and biochemical properties. *Biochem. Pharmacol.* 32:819–826.

3. Deininger, M., C.L. Szumlanski, D.M. Otterness, J. Van Loon, W. Ferber, and R.M. Weinshilboum. 1994. Purine substrates for human thiopurine methyltransferase. *Biochem. Pharmacol.* 48:2135–2138.

4. Paterson, A.R.P., and D.M. Tidd. 1975. 6-Thiopurines. *In* Antineoplastic and Immunosuppressive Agents II. A.C. Sartorelli and D.G. Johns, editors, Springer Verlag New York Inc., New York. 384–403.

5. Weinshilboum, R.M., and S.L. Sladek. 1980. Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity. *Am. J. Hum. Genet.* 32:651–662.

6. Woodson, L.C., J.H. Dunnette, and R.M. Weinshilboum. 1982. Pharmacogenetics of human thiopurine methyltransferase: kidney-erythrocyte correlation and immunotitration studies. *J. Pharmacol. Exp. Ther.* 222:174–181.

7. Weinshilboum, R. 1989. Methyltransferase pharmacogenetics. *Pharmacol. Ther.* 43:77–90.

8. Weinshilboum, R.M. 1992. Methylation pharmacogenetics: thiopurine methyltransferase as a model system. *Xenobiotica.* 22:1055–1071.

9. Krynetski, E.Y., H.-L. Tai, C.R. Yates, M.Y. Fessing, T. Loennechen, J.D. Schuetz, M.V. Relling, and W.E. Evans. 1996. Genetic polymorphism of thiopurine methyltransferase: clinical importance and molecular mechanisms. *Pharmacogenetics.* 6:279–290.

10. Lennard, L., J.A. Van Loon, J.S. Lilleyman, and R.M. Weinshilboum. 1987. Thiopurine pharmacogenetics in leukemia: correlation of erythrocyte thiopurine methyltransferase activity and 6-thioguanine nucleotide concentrations. *Clin. Pharmacol. Ther.* 41:18–25.

11. Lennard, L., J.A. Van Loon, and R.M. Weinshilboum. 1989. Pharmacogenetics of acute azathioprine toxicity: relationship to thiopurine methyltransferase genetic polymorphism. *Clin. Pharmacol. Ther.* 46:149–154.

12. Lennard, L., J.S. Lilleyman, J. Van Loon, and R.M. Weinshilboum. 1990. Genetic variation in response to 6-mercaptopurine for childhood acute lymphoblastic leukaemia. *Lancet.* 336:225–229.

13. Lennard, L., B.E.S. Gibson, T. Nicole, and J.S. Lilleyman. 1993. Congenital thiopurine methyltransferase deficiency and 6-mercaptopurine toxicity during treatment for acute lymphoblastic leukemia. *Arch. Dis. Child.* 69:577–579.

14. Evans, W.E., M. Horner, Y.Q. Chu, D. Kalwinsky, and W.M. Roberts. 1991. Altered mercaptopurine metabolism, toxic effects and dosage requirement in a thiopurine methyltransferase-deficient child with acute lymphoblastic leukemia. *J. Pediatr.* 119:985–989.

15. Schutz, E., J. Gummert, F. Mohr, and M. Oellerich. 1993. Azathioprine-induced myelosuppression in thiopurine methyltransferase deficient heart transplant recipient. *Lancet.* 341:436.

16. Lilleyman, J.S., and L. Lennard. 1994. Mercaptopurine metabolism and risk of relapse in childhood lymphoblastic leukaemia. *Lancet.* 343:1188–1190.

17. Laboratory Medicine Bulletin, Mayo Foundation. Dec. 6, 1991. Thiopurine methyltransferase measurement. 14:1.

18. Van Loon, J.A., and R.M. Weinshilboum. 1982. Thiopurine methyltransferase biochemical genetics: human lymphocyte activity. *Biochem. Genet.* 20:637–658.

19. Szumlanski, C.L., R. Honchel, M.C. Scott, and R.M. Weinshilboum. 1992. Human liver thiopurine methyltransferase pharmacogenetics: biochemical properties, liver-erythrocyte correlation and presence of isozymes. *Pharmacogenetics.* 2:148–159.

20. Honchel, R., I. Aksoy, C. Szumlanski, T.C. Wood, D.M. Otterness, E. Wieben, and R.M. Weinshilboum. 1993. Human thiopurine methyltransferase: molecular cloning and expression of T84 colon carcinoma cell cDNA. *Mol. Pharmacol.* 43:878–887.

21. Lee, D., C. Szumlanski, J. Houtman, R. Honchel, K. Rojas, J. Overhauser, E.D. Wieben, and R.M. Weinshilboum. 1995. Thiopurine methyltransferase pharmacogenetics: cloning of human liver cDNA and presence of a processed pseudogene on human chromosome 18q21.1. *Drug Metab. Dispos.* 23: 398–405.

22. Szumlanski, C., D. Otterness, C. Her, D. Lee, B. Brandriff, D. Kelsell, N. Spurr, L. Lennard, E. Wieben, and R. Weinshilboum. 1996. Thiopurine methyltransferase pharmacogenetics: human gene cloning and characterization of a common polymorphism. *DNA Cell Biol.* 15:17–30.

23. Krynetski, E.Y., J.D. Schuetz, A.J. Galpin, C.-H. Pui, M.V. Relling, and W.E. Evans. 1995. A single point mutation leading to loss of catalytic activity in human thiopurine methyltransferase. *Proc. Natl. Acad. Sci. USA.* 92:949–953.

24. Tai, H.-L., E.Y. Krynetski, C.R. Yates, T. Loennechen, M.Y. Fessing, N.F. Krynetskaia, and W.E. Evans. 1996. Thiopurine *S*-methyltransferase deficiency: two nucleotide transitions define the most prevalent mutant allele associated with loss of catalytic activity in Caucasians. *Am. J. Hum. Genet.* 58:694–702.

25. Otterness, D., C. Szumlanski, L. Lennard, B. Klemetsdal, J. Aarbakke, J.O. Park-Hah, H. Iven, K. Schmiegelow, E. Branum, J. O'Brien, and R. Weinshilboum. 1997. Human thiopurine methyltransferase pharmacogenetics: gene sequence polymorphisms. *Clin. Pharmacol. Ther.* 62:60–73.

26. Yates, C.R., E.Y. Krynetski, T. Loennechen, M.Y. Fessing, H.-L. Tai, C.-H. Pui, M.V. Relling, and W.E. Evans. 1997. Molecular diagnosis of thiopurine S-methyltransferase deficiency: genetic basis for azathioprine and mercaptopurine intolerance. *Ann. Intern. Med.* 126:608–614.

27. Weinshilboum, R.M., F.A. Raymond, and P.A. Pazmiño. 1978. Human erythrocyte thiopurine methyltransferase: radiochemical microassay and biochemical properties. *Clin. Chim. Acta.* 85:323–333.

28. Chadwick, R.B., M.P. Conrad, M.D. McGinnis, L. Johnston-Dow, S.L. Spurgeon, and M.N. Kronick. 1996. Heterozygote and mutation detection by direct automated fluorescent DNA sequencing using a mutant *Taq* DNA polymerase. *Biotechniques.* 20:676–683.

29. Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–517.

30. Nakai, K., and H. Sakamoto. 1994. Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene.* 141:171–177.

31. Mount, S.M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10:459–472.

32. Cooper, D.N., and M. Krawczak. 1996. Human gene mutation database. *Hum. Genet.* 98:629.

33. Maquat, L.E. 1995. When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. *RNA.* 1:453–465.

34. Surdej, P., A. Riedl, and M. Jacobs-Lorena. 1994. Regulation of mRNA stability in development. *Annu. Rev. Genet.* 28:263–282.

35. Naylor, J.A., P.M. Green, A.J. Montandon, C.R. Rizza, and F. Giannelli. 1991. Detection of three novel mutations in two haemophilia A patients by rapid screening of whole essential region of factor VIII gene. *Lancet.* 337: 635–639.

36. McManus, J.F., C.G. Begley, and S. Ratnaike. 1994. Complex pattern of alternative splicing in the normal uroporphyrinogen decarboxylase gene: implications for diagnosis of familial porphyria cutanea tarda. *Clin. Chem.* 40:1884–1889.

37. Berg, L-P., J.M. Soria, C.J. Formstone, M. Morell, V.V. Kakker, X. Estivill, N. Sala, and D.N. Cooper. 1996. Aberrant RNA splicing of the protein C and protein S genes in healthy individuals. *Blood Coagul. Fibrinolysis.* 7:625–631.

38. Krawczak, M., J. Reiss, and D.N. Cooper. 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.* 90:41–54.