Comprehensive study of thiopurine methyltransferase genotype, phenotype, and genotype-phenotype discrepancies in Sweden

Anna Zimdahl Kahlin, Sara Helander, Karin Skoglund, Peter Söderkvist, Lars-Göran Mårtensson and Malin Lindqvist Appell

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-158047

N.B.: When citing this work, cite the original publication.
https://doi.org/10.1016/j.bcp.2019.04.020

Original publication available at:
https://doi.org/10.1016/j.bcp.2019.04.020

Copyright: Elsevier (12 months)
http://www.elsevier.com/
Comprehensive study of thiopurine methyltransferase genotype, phenotype, and genotype-phenotype discrepancies in Sweden

Anna Zimdahl Kahlin¹, Sara Helander¹, Karin Sköglund¹, Peter Söderkvist², Lars-Göran Mårtensson³, Malin Lindqvist Appell¹*

Affiliations

¹Division of Drug Research, Department of Medical and Health Sciences, Linköping University, SE-581 83 Linköping, Sweden.
²Division of Cell Biology, Department of Clinical and Experimental Medicine, Linköping University, SE-581 83 Linköping, Sweden.
³Division of Chemistry, Department of Physics, Chemistry and Biology, Linköping University, SE-581 83 Linköping, Sweden.

*Corresponding author:
Malin Lindqvist Appell, Division of Drug Research, Department of Medical and Health Sciences, Linköping University, SE-581 83 Linköping, Sweden
Telephone: +46 13 286880
Fax: +46 13 149106
E-mail: malin.lindqvist.appell@liu.se

Category; Pharmacokinetics and Drug Metabolism
Abstract

Thiopurines are widely used in the treatment of leukemia and inflammatory bowel diseases. Thiopurine metabolism varies among individuals because of differences in the polymorphic enzyme thiopurine methyltransferase (TPMT, EC 2.1.1.67), and to avoid severe adverse reactions caused by incorrect dosing it is recommended that the patient’s TPMT status be determined before the start of thiopurine treatment. This study describes the concordance between genotyping for common TPMT alleles and phenotyping in a Swedish cohort of 12,663 patients sampled before or during thiopurine treatment. The concordance between TPMT genotype and enzyme activity was 94.5%. Compared to the genotype, the first measurement of TPMT enzyme activity was lower than expected for 4.6% of the patients. Sequencing of all coding regions of the TPMT gene in genotype/phenotype discrepant individuals led to the identification of rare and novel TPMT alleles. Fifteen individuals (0.1%) with rare or novel genotypes were identified, and three TPMT alleles (TPMT*42, *43, and *44) are characterized here for the first time. These 15 patients would not have been detected as carrying a deviating TPMT genotype if only genotyping of the most common TPMT variants had been performed. This study highlights the benefit of combining TPMT genotype and phenotype determination in routine testing. More accurate dose recommendations can be made, which might decrease the number of adverse reactions and treatment failures during thiopurine treatment.

Key words (< 6 st):

Thiopurine, TPMT, pharmacogenetics, genotyping, individualization
1. Introduction

The thiopurine drugs azathioprine (AZA), mercaptopurine (6-MP), and thioguanine (TG) are used for the treatment of childhood leukemia and as immunosuppressants in inflammatory bowel diseases (IBD) and rheumatoid arthritis [1, 2]. Thiopurines undergo a complex metabolism to form metabolites with importance for their mechanism of action, including thioguanine nucleotides (TGNs) and methylated thioinosine monophosphate (meTIMP). TGNs are incorporated into DNA and RNA and inhibit cellular proliferation. One of the key metabolic enzymes is thiopurine methyltransferase (TPMT, EC 2.1.1.67), which catalyzes the transfer of methyl groups from the cofactor S-adenosyl methionine (SAM). TPMT converts the thiopurine drug 6-MP into the inactive metabolite methylated mercaptopurine (MMP) and is also responsible for the last step in the formation of meTIMP, which inhibits de novo purine synthesis and leads to the lack of endogenous nucleotides needed for replication. The thiopurine metabolism has been reviewed previously [2-4].

TPMT is a polymorphic enzyme, 46 allelic variants of the TPMT gene have been identified and reported to the TPMT nomenclature committee and named according to the *allele nomenclature (www.imh.liu.se/tpmtalleles) [5]. Many of these alleles carry SNPs leading to amino acid changes and thus to reduced enzyme activity. Approximately 10% of Caucasian populations have an intermediate level of TPMT activity caused by one defective TPMT allele, while 1 in 300 individuals have two defective alleles and thus have almost undetectable TPMT activity [6]. Three TPMT alleles (*2, *3A, and *3C) are the most common among
individuals with decreased activity and account for 80–95% of the defective alleles [3, 7]. However, there are also allele frequency variations among ethnicities. The frequency of individuals carrying a genetic TPMT variant is lower (2–4%) in Asian populations. In Caucasian and South-west Asian populations, *3A is the most common variant allele, whereas *3C is the most common occurring allele in South-east Asian, African, and Afro-American populations [2, 7, 8].

Low TPMT activity leads to high amounts of active TGNs and an increased risk for myelosuppression and leukopenia, which is the most common severe dose-related adverse event of thiopurines [1, 9]. The Clinical Pharmacogenetics Implementation Consortium (CPIC®) is an international consortium created to facilitate the clinical implementation of pharmacogenetic tests by publishing detailed gene/drug clinical practice guidelines. In the 2018 update of CPIC’s guideline for thiopurine dosing based on TPMT, it is stated that dose adjustment according to TPMT genotype before thiopurine treatment reduces adverse reactions without compromising the therapeutic effect [10]. A high concordance between TPMT genotype and phenotype has been reported [10-12] and has led to the replacement of the phenotyping strategy with genotyping methods that are faster, cheaper, and less labor intensive.

In this paper, the frequencies and concordance between TPMT genotype and phenotype in a large Swedish consecutive cohort of 12,663 individuals is reported. Also, the strategy of extended investigation when a discrepancy between genotype and phenotype was identified is described, as exemplified by the identification of several rare TPMT alleles and the first
characterization of three TPMT alleles (TPMT*42, TPMT*43, and TPMT*44) for which the reduction of enzyme activity is caused by different molecular mechanisms.
2. Material and Methods

2.1 Collection of data

In 2000, the Clinical Pharmacology division, Department of Medicine and Health Sciences, Linköping University, became the first laboratory in Sweden offering clinical laboratory TPMT phenotype testing. In 2004, a genotyping methodology using pyrosequencing for detection of the three most common SNPs was implemented, in addition to phenotyping. The laboratory functioned as a medical advice service for physicians in their decision to adjust and individualize their patients’ doses of thiopurines in order to achieve optimal treatment efficiency. All genotyping and phenotyping data from 2000 to 2015 were collected in a database, the database contains information from 15,975 individuals analyzed in the TPMT routine analysis service during the years 2000–2015 with information regarding diagnosis (based on information from the letter of referral), TPMT genotype, and TPMT phenotype. In many cases repeated measurements of TPMT activity have been performed at different occasions during treatment. Because the data were collected primarily for clinical indications, the population was not randomly selected. Of the 15,975 samples, 71.0% were referred from gastroenterological clinics (comprising patients diagnosed with Crohn’s disease, ulcerative colitis, and hepatitis). Around 14.0% were referred from either dermatology or rheumatology clinics. Hematology clinics referred 6.0% of the samples (including patients diagnosed with acute lymphoblastic leukemia (ALL)). Samples referred from other clinics accounted for 2.3% and included transplantation clinics and patients diagnosed with kidney diseases and vasculitis. The remaining samples (6.7%) were, for example, sent via clinical chemical
laboratories. A part of the data from the database have previously been communicated by Hindorf et al. [13] and Wennerstrand et al. [14]. In this study, the material has been extended to include data from all sampled individuals until the year 2015 (a total of 15,975 individuals).

2.2 Definition of TPMT discrepancies

TPMT enzyme activity was determined in isolated red blood cells as described in section 2.4.1. The enzyme activity was expressed as U/ml packed red blood cells (pRBC) where 1 unit (U) represents the formation of 1 nmol MMP per ml pRBC per hour of incubation [15]. As defined by Pettersson et al. [15], the reference values of normal TPMT enzyme activity were >9.0 U/ml pRBC, for intermediate activity were 2.5–8.9 U/mL pRBC, and for low activity were <2.5 U/ml pRBC. TPMT genotyping for the three most common alleles (TPMT*2, *3A, *3B and *3C) was determined as described in section 2.4.2. An individual was defined as TPMT discrepant if the genotype and phenotype results did not correlate.

2.3 Further investigated patients with discrepancies between genotype and phenotype

2.3.1 Case 1

A man born in 1966 and diagnosed with Crohn’s disease who was prescribed azathioprine without a preceding TPMT status determination was followed for 3 months without any
evident side effects. The patient suddenly became leucopenic and received blood transfusions and intravenous antibiotics. TPMT status was requested and repeated measurements of enzyme activity of 0.5 and 0.6 U/mL pRBC and the genotype TPMT*1/*3A were reported. Based on the discrepancy between genotype and phenotype, new blood samples were requested. The patient also gave consent to request samples from family members. Samples were collected from the patient’s mother, half-brother, wife, and the patient’s two children.

2.3.2 Case 2

Case 2 was a man born in 1989 and diagnosed with systemic lupus erythematosus. The patient’s TPMT enzyme activity was determined to be 7.5 and 8.5 U/mL pRBC at two consecutive measurements, and none of the common defective TPMT alleles were found by routine genotyping. The patient declined further sampling in the study.

2.3.3 Case 3

In a woman born in 1987 and diagnosed with ALL, TPMT status was determined before the start of treatment. The enzyme activity was 2.0 and 2.5 U/mL pRBC at repeated measurements, and none of the common polymorphisms were detected by routine genotyping. New blood samples to obtain RNA were requested, and the patient also gave consent to request samples from family members. Samples were collected from the patient’s mother and
three half-sisters.

2.4 Methods for clinical TPMT laboratory measurements

2.4.1 TPMT enzyme activity

The TPMT enzyme activity in isolated RBC was analyzed as described previously by detecting the amount of $[^{14}C]$-MMP converted from 6-MP and $[^{14}C]$-labeled SAM cofactor as methyl donor after 1 hour of incubation using a scintillation counter [15]. The phenotyping method’s inter assay coefficient of variation (CV) was 4.7% (n = 19, control sample mean 13.5 U/ml pRBC, range 12.7–14.9), and its intra assay CV was 3.3% (n = 8, control sample mean 13.2 U/ml pRBC, range 12.5–13.7).

2.4.2 TPMT genotyping

The presence or absence of the most common defective genetic variants of TPMT – TPMT*2 (rs1800462, c.238G>C p.A80P), TPMT*3A (rs1800460c.460G>A, p.A154T and rs1142345, c.719A>G, p.Y240C), TPMT*3B (rs1800460, c.460G>A, p.A154T), and TPMT*3C (rs1142345, c.719A>G p.Y240C) – were determined by pyrosequencing. DNA was extracted from 200 µl whole blood using the Maxwell DNA purification kit and a Maxwell 16 robot (Promega, Madison, US). The target sequences in exons 5, 7, and 10 were amplified. For exon 5, a total volume of 7 µl containing Hot Star Mastermix 2× (Qiagen Hilden, Germany),
1.1 mM MgCl₂, 0.4 µM of primers (Table 1), and 1 µl DNA was used. For exon 7 and 10, a total volume of 7 µl containing Qiagen Multiplex PCR Kit, 1 µl DNA and primers at concentrations of 0.4 µM (7F, 10R), 0.2 µM (7R), and 0.6 µM (10F) were used. All primers were from Invitrogen (Carlsbad, US) and were designed to avoid amplification of the TPMT pseudogene. One primer in each pair was biotinylated (5´) for later purification of single-stranded DNA. The thermal profile used was 95°C for 15 min; 45 cycles of 95°C for 15 s, 55°C for 90 s, and 72°C for 30s; and a final elongation of 72°C for 10 min. Pyrosequencing analyses to determine the genotype in positions c.238G>C, c.460G>A, and c.719A>G (Table 2) was performed according to the manufacturer’s instructions. In short, single-stranded DNA was purified from the 7 µl of PCR product by capturing it on Sepharose beads (Streptavidin Sepharose High Performance, GE Healthcare, Little Chalfont, UK) dissolved in a binding buffer. The beads were consecutively transferred into ethanol (70 %), 0.2 M NaOH, and washing buffer. Thereafter the beads containing the single-stranded DNA were released into an annealing buffer containing 0.3 µM of 238G>C sequencing primer (singleplex detection) or 0.3 µM of 460G>A and 719A>G primers (multiplex detection), heated to 80°C for 2 minutes, and then allowed to cool to room temperature before the presence of SNPs was determined by pyrosequencing (PyroMark Q96 MD and the PyroMark MD 1.0 software, Biotage, Qiagen).

(Reference included for the first time in table 1: [16])

2.5 TPMT exon sequencing
A routine to follow up individuals with discrepancies in the results from routine genotyping and phenotyping was implemented with a request for a new sample for repeated TPMT enzyme activity measurement as an initial measure followed by extended sequencing for remaining discrepancies. To search for rare or novel TPMT variants, Sanger sequencing of coding TPMT exons (exons 3–10, according to the nomenclature of TPMT exons in [17]) was used as previously described using BigDye Terminator v.3.1 (Applied Biosystems, Foster City, US) labeling and a Genetic Analyzer 3500 (Applied Biosystems) [18].

### 2.6 TPMT cDNA sequencing

To investigate altered splicing, but also for confirmation of the presence of SNPs, sequencing of the patients’ TPMT cDNA was performed. Whole blood was collected in PAXgene Blood RNA tubes and stored at −80°C. The tubes were thawed at room temperature, and RNA was extracted according to the instructions using the PAXgene Blood RNA kit (Qiagen, Hilden, Germany). This method was used in cases 1 and 3. Another protocol using the RNeasy Plus Universal kit (Qiagen, Hilden, Germany) was used to isolate RNA from the clinical laboratory sample in case 2 (frozen EDTA whole blood). As recommended by the manufacturer, QIAzol was added at a 1:10 ratio to 500 µl whole blood during the thawing process, and the sample was allowed to reach room temperature without heating. The sample was homogenized using a QIAshredder column, and thereafter the original protocol for the RNeasy Plus Universal Kit was used. After RNA extraction, the RNA concentration was determined using a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, US) and stored at −80°C until further use. High-Capacity cDNA Reverse
Transcription (LifeTechnologies, Carlsbad, US) was used for cDNA synthesis from 1 µg RNA in a total volume of 25 µl. A protocol for PCR amplification and sequencing, originally described by Lee et al. [19], was used. In short, the TPMT cDNA sequence was amplified using 3 µl cDNA in a total reaction volume of 50 µl HotStarTaq PCR reagents (Qiagen, Hilden, Germany) together with primers for cDNA TPMT exons 3–8 or exons 5–10 that were designed to avoid amplification of the TPMT pseudogene at chromosome 18. The resulting product was first analyzed by agarose gel electrophoresis (E-Gel™ General Purpose Agarose Gels, 2%, Invitrogen) to detect the product’s size. Altered splicing causes several bands of deviating sizes on the gel. A control PCR sample from an individual with TPMT*1/*1 (wildtype) genotype was added to the gel. If several bands occurred on the gel, each DNA product was extracted from the gel using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). The product was labeled using BigDye Terminator v.3.1 (Applied Biosystems) and sequenced as described above. Additional primers were designed in case 2 to include the boundaries of exon 5 (primer forward 5´-CAA GTG GGT GAA CGG CAA G-3´ and reverse 5´-ACA ATG AAA TGT TCC CCG AAG A-3´).

2.7 Recombinant protein expression and purification

The vector pET-28a containing the constructs of human TPMT*1 or TPMT*44 (p.Y166C) was transformed into E.Coli BL21(DE3) CodonPlus cells (Agilent, Santa Clara, US), and TPMT protein was expressed and purified as previously described [18]. In short, cells were cultured in kanamycin-chloramphenicol–enriched media, the expression of TPMT was induced by the addition of isopropyl-β-D-1-thiogalactopyranoside, and expressed over night
at 21°. TPMT proteins were purified using Ni-NTA Superflow beads (Qiagen, Hilden, Germany). The His-tags were cleaved with by biotinylated thrombin (Merck, Darmstadt, Germany), and TPMT proteins were further purified using streptavidin beads (Merck, Darmstadt, Germany), Ni-NTA beads, and gel filtration.

2.8 Thermal denaturation using circular dichroism spectroscopy (CD)

CD measurements were performed as previously described [18]. In short, all spectra were recorded using a Chirascan™ spectrometer (Applied Photophysics, Leatherhead, UK). Thermal denaturation was determined following the CD signal at 222 nm for α-helical unfolding using 3 µM recombinant TPMT protein in a 4 mm cuvette and a temperature range of 5–79°C. The measurements were repeated three times and performed in the presence of a 10-fold molar excess of the cofactor SAM. The results were processed using the computer software CDpal [20]. A Far-UV CD spectra in the range of 215-260 nm was performed to evaluate global secondary structure of TPMTwt and TPMT*44 (p.Y166C) (data not shown).

2.9 Statistics

Descriptive statistics and graphs from the database material were produced using Microsoft Excel 2013 (Microsoft, Redmond, US) and GraphPad Prism 6.03 (GraphPad Software, San Diego, US). The Clopper-Pearson method was used to calculate 95% confidence intervals
(CIs) of conditional proportions and concordance rates. The correlation between TPMT enzyme activity and the factors age, gender and genotype were analyzed using a general linear model. The difference in melting point ($T_m$) between recombinant TPMTwt and TPMT*44 was evaluated using an unpaired t-test. A p value of less than 0.05 was considered statistically significant.

### 2.10 Ethics

The studies were approved by the local ethics committee in Linköping (EPN dnr 01-016 and 2014 194-31). All participants gave their written consent before sampling.
3. Results

3.1 Frequency distribution of TPMT enzyme activity

The database contained data from 15,968 unique individuals (50.8% female and 49.2% male) with at least one phenotype value and complete data. The TPMT activity showed a trimodal distribution (Fig. 1A) with overlaps between groups (Fig. 1B). The TPMT enzyme activity ranged from 0 to 40.8 U/ml pRBC, and 88% had normal TPMT enzyme activity (≥9 U/ml pRBC), 11.6% had intermediate activity (2.5–8.9 U/ml pRBC), and 0.4% had low activity (<2.5 U/ml pRBC). If more than one value was available for the same patient, the most recent TPMT value was chosen for graphical visualization. Seven individuals were excluded because of incomplete data.

3.2 Concordance between TPMT genotype and phenotype

A total of 12,663 patients had both their TPMT phenotype and genotype determined between 2004 and 2015. A comparison was made between the results from genotyping using the routine genotyping for the most common TPMT alleles and the first measured phenotype values. The routine genotyping resulted in 90.4% wildtype, 9.2% heterozygous, and 0.4% homozygous TPMT variants, and the phenotyping method found 86.6% of the individuals
having normal, 12.9% intermediate, and 0.5% low TPMT enzyme activity. When combining
the results from the two methods, the concordance rate was 94.5% (95% CI; 94.1–94.9)
between genotyping and phenotyping. More specifically, the genotyping method identified
99.1% (95% CI; 98.9–99.2) of all individuals with normal TPMT enzyme activity, 64.4%
(95% CI; 62.0–66.7) with intermediate TPMT enzyme activity, and 83.6% (95% CI; 71.9–
91.8) with poor TPMT enzyme activity (Fig. 2). According to the genotyping result, 588
individuals (4.6%) had lower and 107 (0.8%) had higher TPMT enzyme activity than
expected. The most commonly occurring TPMT discrepancy was when both alleles were
detected as TPMT*1 by routine genotyping but the TPMT enzyme activity was in the
intermediate TPMT activity range (4.6%, n = 578).

3.3 Extended sequencing of individuals with discrepancies between TPMT genotype and
phenotype

Out of 11,433 individuals with genotype TPMT*1/*1, 579 (4.6%) had a TPMT enzyme
activity below 9.0 U/ml pRBC at the first measurement (median 8.1 U/ml pRBC, median
range 2.0–8.9). For all individuals with lower activity than expected, a new blood sample for
repeated enzyme activity measurement was requested, 310 new samples were received and
228 (73.5%) had results of 9 U/ml pRBC or higher at this second measurement. Among the
remaining discrepancies between routine genotyping and phenotyping results, 10 individuals
previously detected as TPMT*1/*1 were identified as heterozygous carrying one additional
rare or novel defective allele after sequencing all of the coding TPMT exons (Table 3). Five
individuals detected as heterozygous for one of the most common TPMT variants were
defined as totally deficient after identifying additional non-functional alleles. Taking into consideration the frequencies of discovered genotypes after extended sequencing, 90.3% were TPMT wildtype (*1/*1), 9.2% were heterozygous (*1/*variant), and 0.5% were homozygous deficient (*variant/*variant). In total, 15 individuals (0.1%) with rare or novel TPMT alleles not detected in the routine genotyping method were identified. Some of these alleles had previously been reported from our research group [21-23]. When describing the genotype frequency, no data from known or requested family members were included in order to avoid artificially high frequencies of rare genotypes. Individuals genotyped as TPMT*1/*1 (n = 11,433) had a TPMT enzyme activity range of 3.3–40.8 U/ml pRBC, heterozygous individuals 0–12.7 U/ml pRBC (n = 1170), and individuals carrying two defect alleles 0–5.5 U/ml pRBC (n = 60) (Fig. 1B). After further investigation and repeated activity measurements, the overall concordance rate between genotype and phenotype was 96.4% (95% CI; 96.1–96.8).

3.4 Correlation between TPMT enzyme activity, genotype, age and gender

TPMT enzyme activity’s correlation to the factors age, gender and genotype were analyzed using a general linear model, Table 4. Individuals with complete data regarding TPMT enzyme activity (most recent value), genotype, birth date, sample date and gender were included in the analysis, resulting in 12,654 included individuals (mean±SD; 39±22 years). The interaction between gender and genotype was removed from the model because of no significance. Significant factors were gender and genotype and there were also significant interactions between age and both gender and genotype. This means that the level of TPMT enzyme activity differed between genotypes and also between males and females.
Additionally, the age was not affecting the TPMT enzyme activity overall. However, when dividing the data according to gender or genotype there was an interaction with age. The TPMT enzyme activity was higher in males than in females, Fig. 3. Additionally, the enzyme activity was higher in older individuals than in younger, in both wildtypes and heterozygous.

3.5 Discrepant results in samples from hematology patients

A well-known phenomenon when measuring TPMT enzyme activity is false low activity in the beginning of ALL treatment [14, 24] due to disturbed hematopoiesis. In individuals with a hematology diagnosis (n = 798), the overall concordance rate between routine genotyping and the first phenotype measurement was 70.7% (95% CI; 67.4–73.8). The majority of the intermediate metabolizers (78.7%, 95% CI; 73.5–83.3, n = 225) were classified as TPMT*1/*1 using the routine genotyping method.

3.6 Case 1 – the TPMT*42 allele carries an insertion in the TPMT gene

Samples from the patient and his family members were analyzed (Fig. 4). The patient’s half-brother and son carried one TPMT*3A allele and had intermediate TPMT enzyme activities of 6.9 and 7.5 U/ml pRBC, respectively. The mother and the daughter of the patient had intermediate enzyme activities of 8.4 and 7.7 U/ml pRBC, respectively. The patient’s wife had normal enzyme activity (13.1 U/ml pRBC) and the genotype TPMT*1/*1 (Fig. 4A). Sequencing results from the patient, his mother, and his daughter in both genomic DNA and
cDNA showed an insertion of an A nucleotide at position c.95_96 in TPMT exon 3 (Fig. 4B, 4C). The insertion caused a silent amino acid shift in p.K32, AAG shifted to AAA (Fig. 4D). The insertion caused a frameshift that led to a nonsense sequence of amino acids, and a premature termination codon was created at amino acid position 58 causing an incomplete TPMT protein. The allele containing the insertion was named TPMT*42 (c.95_96insA, p.K32KfsX58, ClinVar accession number SCV000778513, rs759836180) by the TPMT nomenclature committee.

3.7 Case 2 – Two SNPs lead to deletion of exon 5 (TPMT*43)

A clinical laboratory blood sample from the patient was analyzed (Fig. 5). The sequencing of genomic DNA showed two genetic variants located in the first base (c.234G>T) and 29th base (c.262G>A) of exon 5 (Fig. 5A). When visualizing the cDNA PCR products on an agarose gel before sequencing, two extra bands occurred on the gel indicating an altered splicing pattern (Fig. 5B). The bands were extracted and sequenced separately. The two longest bands showed the wildtype TPMT sequence with exon 5 included (band 1, 298 bp, Fig. 5C) and the TPMT sequence without exon 5 (band 2, 166 bp, Fig. 5D). It was not possible to read any sequence from the sequencing result of the smallest band from the gel. The intermediate enzyme activity and the fact that one band contained the sequence without exon 5 and the other band contained the wild type sequence proves that both genetic variations are located on the same allele and it was therefore named TPMT*43 (c.234G>T, p.W78C, ClinVar accession number SCV000778514 and c.262G>A, p.G88S, ClinVar accession number SCV000778515, rs753545734) by the TPMT nomenclature committee. In conclusion, the TPMT*43 allele
caused a nonfunctional TPMT enzyme with a deletion covering the entire exon 5.

3.8 Case 3 – the TPMT*44 allele causes an amino acid substitution in the TPMT protein

Samples from the patient and her family were analyzed (Fig. 6). Sequencing of TPMT exons 3–10 was performed using genomic DNA and cDNA (Fig. 6A, 6B). A previously uncharacterized heterozygous sequence variant, c.497A>G, which caused the amino acid change p.Y166C, was identified in the third base of TPMT exon 8. The sequencing allele was named TPMT*44 (c.497A>G p.Y166C, rs201695576, ClinVar accession number SCV000778516) by the TPMT nomenclature committee. The nucleotide substitution was found heterozygously in the patient, her mother, and one of her half-sisters (Fig. 6C). The two other half-sisters had no deviating sequence after sequencing TPMT exons 3–10. This agrees with the phenotype results showing that the patient’s mother and the half-sister carrying the sequencing variant also had intermediate TPMT enzyme activity (4.4 and 8.7 U/ml pRBC, respectively) and one half-sister had normal enzyme activity (9.9 U/ml pRBC). Interestingly, the third half-sister with the wild-type TPMT*1/*1 genotype displayed a TPMT enzyme activity of 6.7 U/ml pRBC, which is in the intermediate activity range.

To further investigate the consequence of the amino acid substitution, recombinant TPMT with the mutation p.Y166C was produced, and the stability of the resulting protein was determined using thermal denaturation (Fig. 6D). The thermal midpoint of denaturation (T_m) of TPMT wildtype protein in the presence of SAM cofactor was 49.4 ± 0.3°C, and the T_m of
TPMT*44 (p.Y166C) was clearly decreased to 36.6 ± 0.4°C (df = 5, p < 0.0001). The overall secondary structure, evaluated using CD, was unchanged (data not shown).
4. Discussion

This study describes a large cohort of patients including information about diagnosis and TPMT phenotype and genotype before and/or during thiopurine treatment. The cohort was studied after the first TPMT measurement and clinical genotyping for the three most common TPMT SNPs as well as after extended sequencing and repeated activity measurements. The concordance rate between the genotypes and phenotypes of TPMT was high (94.5% before and 96.4% after extended investigation, n = 12,663). Previously reported TPMT genotype and phenotype concordance rates have been summarized by Schaeffeler et al. [12] and range between 76% and 100% (n = 21–1,214 individuals). There are differences in how well the methods identify the three different levels of metabolizers. In our study 99.1% of the normal metabolizers were found by the routine genotyping method, but there was a considerably lower concordance rate in the group of intermediate metabolizers (64.4%), similar to other studies [12]. In our comparison we included 12,663 individuals, which is, to the knowledge of the authors, so far the largest genotype-phenotype correlation study of TPMT.

Under some circumstances, genotyping is to be preferred because it is known that the TPMT enzyme activity is artificially low because of the treatment (e.g. after blood transfusions [9]) or because of the disease itself. In the early treatment of ALL, an artificially low TPMT enzyme activity is common due to disturbed hematopoiesis and the excess of old RBC that display a lower enzyme activity than young RBC [14, 24]. In our study we showed that most of the patients (78.7%) diagnosed with hematological disease and having intermediate
enzyme activity were classified as TPMT*1/*1 by the routine genotyping method, and this phenomenon was also seen in case 3.

In our material, after the first activity measurements 0.8% displayed a higher and 4.6% displayed a lower enzyme activity than expected according to their genotype. On the one hand, the statistically most reliable results are from the genotyping methods (this study and [12, 13]) regarding classification into wildtype, heterozygous defective, and homozygous defective, and it is also the most cost-effective method. On the other hand, in the clinical context, the most clinically relevant results come from the TPMT enzyme activity measurements, which determine the actual metabolic function of TPMT in the body. The main metabolic site of thiopurines is in the liver [26], and their main site of action is believed to be in white blood cells (DNA incorporation and immunosuppression). However, the consensus of TPMT enzyme activity measurements in RBC, liver or kidney tissue, or leukemic blasts was established during the 1990s [27, 28], and in routine clinical practice TPMT enzyme activity is determined in isolated RBC [15] or in whole blood [29]. The results from measurements in whole blood are said to be less variable due to an easier laboratory procedure. The TPMT activity could also be affected by inhibitors that are removed when measuring activity in isolated RBC. Recently, a study of 124 human liver samples failed to show a correlation between hepatic TPMT activity, mRNA expression, and TPMT protein levels [26] supporting the theory that the enzymatic activity is affected by posttranslational modifications and changed protein degradation pattern. After repeated TPMT enzyme activity measurements, many of the discrepant results in our study were within the expected enzyme activity range, but a second sample was not always obtained from the clinics. Some discrepant cases could possibly be explained by a result close to the cut off level of next reference
In clinical laboratories analyzing TPMT, the most common way to analyze the genotype is to only genotype for the three most common TPMT sequence variants (c.238G>C, c.460G>A, and c.719A>G) [31]. With this arrangement, individuals carrying rare or novel TPMT alleles would be classified as carrying a wildtype alleles. An undiscovered defective TPMT allele causing lower TPMT enzyme activity could be devastating for the patient if not followed up by other blood status parameters. To use a strategy where both genotype and phenotype are analyzed would give the best individualized dosage prediction based on TPMT, but this would be expensive and would require high competence from the laboratory. In our study, by using this strategy of combining the methods and adding the extended TPMT sequencing of individuals with remaining discrepancies, 15 individuals (0.1%) carrying rare or novel TPMT alleles were identified.

The analysis of the TPMT enzyme activity’s correlation to gender, age and genotype show some interesting findings. The correlation between gender or age have previously been studied [15, 26, 30-33] with varied results. Our data show that men have higher TPMT enzyme activity than woman in all three classes of genotypes (Fig. 3). The difference between gender appears to be less in younger individuals. We can also see that older individuals have higher TPMT activity than younger and this phenomenon are more pronounced in wildtype individuals and heterozygous males. Speculations why studies differ in terms of the impact of age and gender on TPMT enzyme activity could be smaller sample sizes, misclassification of rare or novel genotypes, sub-grouping or whereas the investigated population are healthy volunteers or not. However in our study, as well as other studies investigating these
correlations, any differences found in TPMT enzyme activity are so small that they are most likely not of clinical importance.

In this study, three alleles were identified and characterized for the first time, including TPMT*42, TPMT*43, and TPMT*44. TPMT*44 (p.Y166C) leads to a single amino acid substitution, which is the most common reason for low TPMT enzyme activity. Previous studies have shown that an amino acid shift can give rise to various effects in the TPMT protein [34, 35]. Amino acid Y166 is part of the Y166-D151-Y180 hydrogen bonding network in the TPMT protein core [36] and is important for the stabilization of the protein, which also is shown by the lower thermal denaturation temperature observed for TPMT*44 compared to TPMTwt in the presence of SAM (Fig. 6D). The observed decreased stability for other SNPs causing amino acid shifts in this region is explained by causing a shortened protein half-life in vivo due to increased degradation, as investigated and discussed by Wennerstrand et al. [36]. It is very likely that the reason for the loss of enzymatic activity in vivo as well as decreased stability during thermal denaturation in vitro for TPMT*44 is due to the localization of the variant in the hydrogen bonding network.

In TPMT*42, the insertion of an extra A in the DNA sequence coding for the amino acids caused a frameshift and a premature stop codon. TPMT*42 is the first allele named by the TPMT nomenclature committee, with an insertion as a cause of the loss of enzyme activity. In TPMT*43, two genetic variations were found in the first and 29th nucleotides of exon 5. The enzymatic activity was intermediate, and the sequencing result showed a sequence containing the wild type sequence and the other with a deletion of the entire sequence of exon 5. Therefore, the conclusion was that both genetic variations were located on the same allele and
caused a disruption of the splicing site. There are a few reports of proven or suspected deviating splicing patterns caused by SNPs in TPMT. Using next generation sequencing, a deletion of exon 5 was reported together with TPMT*34 (located in exon 5) and the intron mutation c.366+58T>C [37]. However, no phenotype or patient data were described in that study, and the extra band appearing in the gel analysis was not sequenced. Exon skipping was also proposed to be the cause of the low TPMT enzyme activity in a patient carrying TPMT*15, located in the last nucleotide of intron 7 (IVS7-1G>A [22]. The growing numbers of known genetic factors influencing TPMT enzyme activity raises the demands on a potential future method that makes it possible to detect all TPMT variations, including those located in introns and promoter regions causing altered splicing or transcription.

Interestingly, our research group has identified another 50 individuals with intermediate TPMT enzyme activity and a TPMT*1/*1 genotype that have been sequenced (exons 3–10) without observed variations. Additionally, there are a few individuals with an enzyme activity above 9.0 U/ml pRBC together with a heterozygous genotype after sequencing exons 3–10. Many mechanisms for additional regulation of TPMT enzyme activity, in addition to the genotype, have been suggested, for example, levels of the SAM cofactor [38, 39], polymorphisms in the MTHFR and PACSIN2 genes [40, 41], and variable-number tandem repeats [42-44]. In addition, drugs commonly used in these patient cohorts have been proposed to inhibit TPMT activity [14, 45-47]. However, conflicting conclusions have been drawn for many of the proposed mechanisms [48, 49], and recently two large genome-wide association studies [27, 31], including several cohorts, showing that the primary cause of interindividual variability in TPMT enzyme activity is the genetics of TPMT. However, it was evident in the study by Tamm et al. [27] that additional contributory causes could not be
excluded. One could speculate that contributing mechanisms might be more or less important under certain circumstances and thus make these mechanisms difficult to investigate. In addition to the phenomenon of increasing TPMT enzyme activity in the early treatment of ALL, the effect of exposure to thiopurines during 20 weeks treatment in IBD patients have been studied [11], and overall no significant changes in the TPMT activity were observed. To our knowledge the effect of prolonged exposure to thiopurines on TPMT enzyme activity has not been thoroughly investigated.

In summary, this study presents the largest population with high concordance rate between genotyping for the three most common TPMT SNPs and phenotyping for TPMT activity. By extended sequencing of TPMT exons in case of discrepancies in the results, 15 individuals carrying rare or novel TPMT alleles were found. We have characterized the three alleles TPMT*42, *43, and *44 found in this study. Genotyping and phenotyping methods are beneficial for different patient groups, which justifies the use of a combination of genotyping and phenotyping in order to detect as many individuals as possible who are at risk of treatment failures and adverse reactions during thiopurine treatment.

Acknowledgments

This study was generously supported with grants from The Swedish Children´s Cancer Foundation (MLA), The Medical Research Council of Southeast Sweden (MLA), The Swedish Society of Medicine Linköping (AZK), and Östgötaregionens cancerfond (AZK). We would also like to thank Curt Peterson, Malin Larsson (NBIS), Lars Valter (Forum Östergötland), Patrik Lundström, and Markus Niklasson for valuable help with bioinformatics.
and statistical analyses and scientific discussions. The skillful technical assistance of Karin Wåhlén and Roshni Narshi has been greatly appreciated.

Conflicts of interest

The authors declare none conflicts of interest.
References


Figure legends

**Fig. 1.** A) The frequency of the TPMT phenotypes in the database of the Clinical Pharmacology Department, Linköping University, containing samples from 15,968 individuals collected before or during thiopurine treatment. If repeated activity results were available, the most recent value was used. B) Distribution of the TPMT phenotypes according to the genotype for all individuals with both genotype and phenotype values available (n = 12,663). The bars corresponding to heterozygous defective (*1/*X) individuals are transparent and located in front of bars corresponding to wildtype (*1/*1) individuals. Extended sequencing results are included in the genotype data, and if repeated activity results were available, the most recent value was used. Individuals with TPMT enzyme activity in the range of 25.1–40.8 U/ml pRBC (n = 18) were excluded from figure 1B.
Fig. 2. Concordance between TPMT genotype and phenotype using the first enzyme activity measurement and the routine genotyping method, n = 12,663. The values in white boxes represent individuals where TPMT enzyme activity and genotype were in concordance, 94.5\% (95% CI: 94.1-94.9), and values in light gray boxes indicate individuals where TPMT enzyme activity deviated from the genotyping result. Cut off levels for TPMT enzyme activity were normal >9 U/ml pRBC, intermediate 2.5–8.9 U/ml pRBC, and low <2.5 U/ml pRBC.
Fig. 3. Regression lines derived from parameter estimates between TPMT enzyme activity and age for males (---) and females (—) in TPMT wildtypes (*1/*1), heterozygotes (*1/*X) and homozygotes (*X/*X) based on the general linear model.
**Fig. 4.** Overview of case 1, TPMT*42. A) The pedigree of the patient (dotted line) and his family with the TPMT*42 allele (striped grey). The results of TPMT enzyme activity and genotype analyses are shown. B) Sequencing histogram from genomic DNA, TPMT exon 3. The insertion is marked with an arrow. C) Sequencing histogram from cDNA, TPMT exon 3. The cDNA position of the insertion c.95_96insA is marked with an arrow. D) Visualization of the amino acid sequence in the TPMT wild type protein (*1) and TPMT p.K32KfsX58 (*42). The insertion of c.95_96insA leads to a frameshift (underlined red) and a premature stop codon in the sequence.
**Fig. 5.** Overview of case 2, TPMT*43. A) Sequencing histogram from genomic DNA, TPMT exon 5. The detected SNPs are marked with arrows. B) Agarose gel of the cDNA PCR product from the patient (P) and control (C). Reference (R) 100–1000 bp. The bands marked with 1 and 2 show the PCR products of allele 1 and 2, respectively. C) Sequencing histogram from cDNA allele 1, exon 5. The cDNA positions of the two detected SNPs in c.234 and c.262 are marked with arrows but were detected as wild type. Disorders in the sequence are located 5–8 bases before the SNP of interest. D) Sequencing histogram from cDNA allele 2 in which exon 5 is missing.
Fig. 6. Case 3, TPMT*44. A) Sequencing histogram illustrating the A>G transition (black arrow) in the genomic DNA located in the third base of TPMT exon 8. B) Sequencing histogram of cDNA with c.497A>G marked with a black arrow. The presence of both exon 7 and 8 indicate that the substitution does not affect the splicing process. C) Pedigree with genotype and phenotype results of the patient (dotted line) and her family. D) The thermal denaturation analysis of the recombinant TPMT protein shows decreased protein stability for TPMT*44 (p.Y166C) compared to TPMTwt. The CD signal monitored at 222 nm from recombinant TPMTwt (●) and TPMT p.Y166C (▲) during increasing temperature result in a $T_m$ of 49.4 ± 0.3°C and 36.6 ± 0.4°C, respectively. The graph shows data from one representative measurement of both proteins.
<table>
<thead>
<tr>
<th>Target</th>
<th>Primer forward</th>
<th>Primer reverse</th>
<th>Fragment size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon V</td>
<td>BiOCCC TCT ATT TAG TCA TTT GAA AAC</td>
<td>ACT TTT GTG GGG ATA TGG AT</td>
<td>180 bp</td>
</tr>
<tr>
<td>Exon VII\textsuperscript{a,b}</td>
<td>TGT TGA AGT ACC AGC ATG CAC</td>
<td>BiOAAA ATT ACT CAT TTG CGA TCA</td>
<td>379 bp</td>
</tr>
<tr>
<td>Exon X\textsuperscript{b}</td>
<td>BiOCCC AGC CAA TTT TGA GTA TT</td>
<td>CAA TTC CTC AAA AAC ATG TCA</td>
<td>495 bp</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Primer sequences previously published in [16]. \textsuperscript{b}Primers used together for multiplex detection. BiO Biotinylated primer.
Table 2
Sequencing primers and settings for pyrosequencing reactions

<table>
<thead>
<tr>
<th>Target position</th>
<th>Sequencing primer (5´-3´)</th>
<th>Sequence to analyze</th>
<th>Dispensation order</th>
</tr>
</thead>
<tbody>
<tr>
<td>238G&gt;C</td>
<td>TGT CCC CGG TCT G</td>
<td>C/G AAA CCT GCA</td>
<td>TCG TAC TGC</td>
</tr>
<tr>
<td>460G&gt;A&lt;sup&gt;a&lt;/sup&gt;</td>
<td>TGA TTT GGG ATA GAG GA</td>
<td>G/A CAT TAG TTG CCA TCA AT</td>
<td>CGA GCT AGT ATC TAG</td>
</tr>
<tr>
<td>719A&gt;G&lt;sup&gt;a&lt;/sup&gt;</td>
<td>CTC ATT TAC TTT TCT GTA AG</td>
<td>TAG A T/C ATA ACT TTT GAA A</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Primers used together for multiplex detection.
Table 3
TPMT genotypes among the individuals registered in the database after investigation of discrepancies.a.

<table>
<thead>
<tr>
<th>TPMT genotype</th>
<th>Number of individuals</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wildtype</td>
<td>*1/*1</td>
<td>11433</td>
</tr>
<tr>
<td>Individuals</td>
<td>*1/*2</td>
<td>17</td>
</tr>
<tr>
<td>carrying one</td>
<td>*1/*3A</td>
<td>1030</td>
</tr>
<tr>
<td>variant allele</td>
<td>*1/*3B</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>*1/*3C</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>*1/*9</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>*1/*14</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>*1/*23</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>*1/*31&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>*1/*43</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>*1/*44</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>Heterozygotes</td>
<td>1170</td>
</tr>
<tr>
<td>Individuals</td>
<td>*3A/*3A</td>
<td>40</td>
</tr>
<tr>
<td>carrying two</td>
<td>*3A/*3C</td>
<td>11</td>
</tr>
<tr>
<td>variant alleles</td>
<td>*3A/*3B</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>*3C/*3C</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>*2/*3A</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>*3A/*14&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>*3A/*15&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>*3C/*23</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>*3A/*23&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>*3C/*42</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>---------</td>
<td>-----</td>
</tr>
<tr>
<td>Homozygotes</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>All individuals</td>
<td>12663</td>
<td></td>
</tr>
</tbody>
</table>

*TPMT discrepancy is defined as TPMT*1/*1 combined with enzyme activity <9 U/ml pRBC or TPMT*1/*X and enzyme activity <2.5 U/ml pRBC after the routine analysis.  
*Results previously reported as TPMT*28 in reference [21].  
*Results previously reported in reference [22].  
*Results previously reported in reference [23].
Table 4. Results from the general linear model

<table>
<thead>
<tr>
<th>Factor</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>0.001</td>
</tr>
<tr>
<td>Genotype</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age</td>
<td>0.184</td>
</tr>
<tr>
<td>Gender * Age</td>
<td>0.001</td>
</tr>
<tr>
<td>Genotype * Age</td>
<td>0.003</td>
</tr>
<tr>
<td>TPMT genotype</td>
<td>TPMT enzyme activity (U/mL pRBC)</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
</tr>
<tr>
<td>Wildtype</td>
<td>10,864 99.1 (98.9-99.2)</td>
</tr>
<tr>
<td>Heterozygous</td>
<td>103 0.9 (0.8-1.1)</td>
</tr>
<tr>
<td>defective</td>
<td></td>
</tr>
<tr>
<td>Homozygous</td>
<td>0</td>
</tr>
<tr>
<td>defective</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10,967 (100%)</td>
</tr>
</tbody>
</table>
gDNA exon III forward

ATGGCAAGACAAAGTGGGTGAACGGCAAGACTGCTTT

ATGGCAAGACAAAGTGGGTGAACGGCAAGACTGCTTT

gDNA exon III forward

ATGGCAAGACAAAGTGGGTGAACGGCAAGACTGCTTT

ATGGCAAGACAAAGTGGGTGAACGGCAAGACTGCTTT

c.95_96insA

TPMT wt protein

TPMT p.K32KfsX58

TPMT wt protein

TPMT p.K32KfsX58


Y F P Stop
A) gDNA intron ←→ exon VIII forward
TGTTCTTTTG TAGCTATGCA GATACAATGTTTT

B) cDNA exon VII ←→ exon VIII forward
TGATCGCAAATGCTATGCA GATACAATGTTTT

497A>G

C) Pedigree

D) Fraction unfolded vs. Temperature (°C)

- TPMT wt
- TPMT p.Y166C