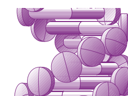
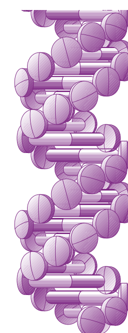
Research Article

For reprint orders, please contact: reprints@futuremedicine.com



Pharmacogenomics



Novel motif of variable number of tandem repeats in *TPMT* promoter region and evolutionary association of variable number of tandem repeats with *TPMT*3* alleles

Dunja Urbančič[1], Alenka Šmid[1], Gabriele Stocco[2], Giuliana Decorti[3,4], Irena Mlinarič-Raščan[1] & Nataša Karas Kuželički*[,1]

[1]Department of Clinical Biochemistry, Faculty of Pharmacy, University of Ljubljana, Aškerčeva cesta 7, 1000 Ljubljana, Slovenia

[2]Department of Life Sciences, University of Trieste, 34127 Trieste, Italy

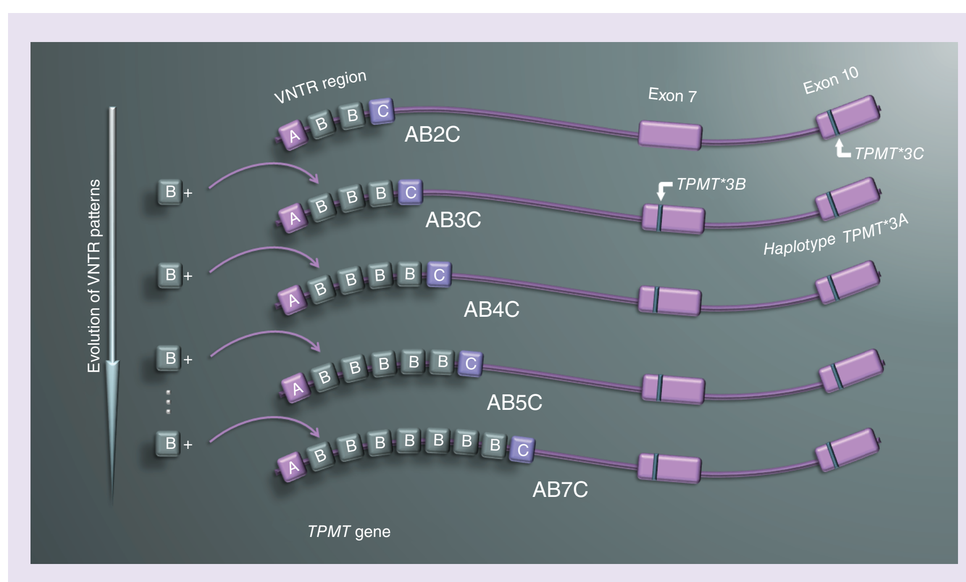
[3]Laboratorio di Diagnostica Avanzata Traslazionale, Institute for Maternal & Child Health - IRCCS 'Burlo Garofolo', 34127 Trieste, Italy

[4]Department of Medicine, Surgery & Health Sciences, University of Trieste, 34127 Trieste, Italy

*Author for correspondence: Tel.: +386 1 4769 629; natasa.karaskuzelicki@ffa.uni-lj.si

Aim: SNPs in the gene for TPMT exemplify one of the most successful translations of pharmacogenomics into clinical practice. This study explains the correlation between common SNPs and variable number of tandem repeats (VNTR) in promoter of the gene. **Materials & methods:** We determined VNTR polymorphisms, as well as *TPMT*2* and *TPMT*3* SNPs and TPMT activity in Slovenian and Italian individuals and lymphoblastoid cell lines. **Results:** We observed a previously unreported VNTR allele, AB7C, in a *TPMT*3A* heterozygous individual. VNTRs with two (AB2C) and three or more (ABnC, $n \geq 3$) B motifs were statistically significant in complete linkage disequilibrium ($D' = 1$, $r^2 = 1$, $p < 0.0001$) with the *TPMT*3C* and *TPMT*3A* alleles, respectively. **Conclusion:** The study provides insights into the stepwise evolution of *TPMT*3* alleles from **3C* to **3A*, with increasing number of B motifs in the VNTR region.

Graphical abstract:



First draft submitted: 23 July 2018; Accepted for publication: 24 September 2018; Published online: 22 October 2018

Future Medicine

Keywords: acute lymphoblastic leukemia • genetic polymorphisms • inflammatory bowel disease • linkage disequilibrium • lymphoblastoid cell lines • thiopurine *S*-methyltransferase • TPMT activity • TPMT • variable number of tandem repeats • VNTR

Due to high correlation of its activity with common genetic polymorphisms, TPMT exemplifies one of the most successful translations of pharmacogenomics into clinical practice [1]. TPMT plays an important part in the deactivation pathway of thiopurine drugs by directing conversion of thiopurines, 6-mercaptopurine (6MP) and 6-thioguanine, to the methylated entities, 6-methyl-MP (6MMP) or 6-methylthioguanine, respectively [2]. This *S*-adenosylmethionine-dependent methylation decreases levels of 6-thioguanine nucleotides, the active metabolites of thiopurine drugs, which are used in patients with acute lymphoblastic leukemia (ALL), transplantation procedures and autoimmune disorders [3–5]. The population distribution of TPMT activity is trimodal [2]. The majority of individuals (89%) have normal TPMT activity, around 11% of population have decreased activity and around 0.3% bears completely deficient enzyme [6]. Often, the decrease in TPMT activity is due to SNPs in the TPMT gene, among which *TPMT*3A*, *TPMT*3C* and *TPMT*2* are the most common [7]. Individual genetic variability of *TPMT* is mirrored in levels of 6-thioguanine nucleotides. When patients with lower enzyme activity are treated with standard thiopurine doses, they exhibit a lower relapse rate, but are more prone to severe hematologic side effects [3,8].

In addition to the variability caused by SNPs in the coding region of *TPMT*, its G/C rich promoter is variable as well, containing DNA motifs in the form of a variable number of tandem repeats (VNTR) [9]. This repetitive DNA region consists of three motifs in its internal structure, named A, B and C (Figure 1) [10]. Reported human VNTR alleles are made up of 1–7 repeats of motif A, 1–6 repeats of motif B and a single motif C. The motifs always appear in the same order: AmBnC [10]. The overall length of repeats ranges from three to nine motifs, which classifies VNTR alleles in the *TPMT* promoter as minisatellites [11]. The evolutionary mechanism, that best describes the formation of novel VNTR alleles in the *TPMT* promoter, is the stepwise mutation model (SMM) [12,13]. The model assumes gradual gaining (or losing) of one single motif at the time, explaining the phylogenetic relationships between VNTR alleles [13].

Early studies attempted to determine the influence of VNTRs in the *TPMT* promoter on TPMT enzyme activity. Lower activity of the enzyme was associated with VNTR genotypes containing more than five repeats in total [9,14,15]. However, several of these studies did not take into account the structure of the VNTR region (i.e., the exact number of A and B motifs), or did not label the alleles consistently, preventing exact conclusions. Recently, the number of A motifs in VNTR has been negatively correlated with the transcription rate of TPMT [16,17]. The results have been replicated in a clinical trial on patients treated with thiopurines [18], suggesting the importance of precise characterization of VNTR patterns in terms of the number of A and B motifs for clinical applications.

The function and the mechanism of action of the VNTR region remain unknown. In this study, we investigated the association of VNTR architecture with TPMT activity and its connection to common clinically relevant genetic polymorphisms in the TPMT gene. We sequenced the *TPMT* promoter, genotyped *TPMT* for *TPMT*2* and *TPMT*3* alleles and measured TPMT activity in healthy Slovenian volunteers. Results were validated in a cohort enriched for variant *TPMT*3* alleles, comprising nonrandomly selected Italian patients with ALL or inflammatory bowel disease (IBD) and lymphoblastoid cell lines (LCL) from healthy Israeli and Estonian donors.

Materials & methods

Blood samples

Venous blood (3–5 ml) from 132 unrelated healthy Caucasian volunteers from Slovenia [6] was used for TPMT genotyping. TPMT activity was measured in the red blood cells (RBC) of 29, randomly chosen, Slovene individuals [6]. Since the frequency of *TPMT*3* alleles in the Slovenian cohort was too low to properly perform the linkage disequilibrium (LD) analysis, randomly selected Slovenian cohort was complemented by the samples of individuals previously determined to be carriers of *TPMT*3* alleles. For VNTR analysis 24 such samples were nonrandomly selected from Italian ALL or IBD patients [19,20].

Lymphoblastoid cell lines

In order to further enrich the study cohort with *TPMT*3* alleles, we conducted experiments on 41 LCLs of consenting healthy adults, 30 of which were obtained from the National Laboratory for the Genetics of Israeli populations, Tel-Aviv University, Israel [21]. LCLs from the National Laboratory for the Genetics of Israeli populations

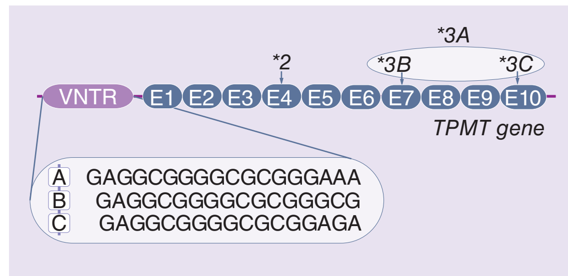


Figure 1. The structure of variable number of tandem repeats in *TPMT*. The sequence of A, B and C VNTR motifs and location of **3* polymorphisms in the coding region of *TPMT* are shown.

Table 1. The list of analyzed SNPs in the *TPMT* coding region, their rs numbers, nucleotide changes and TaqMan Assay IDs.

rs number	Allele symbol	Nucleotide change	TaqMan Assay ID
rs1800462	*TPMT*2*	238G >C	C_12091552_30
rs1800460	*TPMT*3B*	460G >A	C_30634116_20
rs1142345	*TPMT*3C*	719A >G	C_19567_20

were generated from the lymphocytes of 30 healthy donors – 27 Ashkenazi individuals, 2 Palestinian individuals and 1 Ethiopian individual as previously described [21]. The other 11 LCLs were prepared from the lymphocytes of selected Caucasian subjects from the Estonian Biobank, Estonian Genome Center, University of Tartu [22]. The lymphocytes of healthy adult donors were immortalized *in vitro* by Epstein–Barr virus. Cells were cultured in RPMI medium (Merck, Sigma-Aldrich, Darmstadt, Germany), supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, Gibco, MA, USA), 4mM L-glutamine and antibiotics (100 U/ml penicillin; 100 µg/ml streptomycin; 250 ng/ml amphotericin B; all from Merck, Sigma-Aldrich) and kept under humidified conditions at 37°C and 5% CO_2 [21].

Extraction of DNA and genotyping

For blood samples, the extraction of DNA and determination of *TPMT**2 and *TPMT*3* alleles has been described previously [6,19,20].

For LCLs, DNA was extracted from 2.5×10^6 cells according to the manufacturer's instructions using the QIAamp DNA Blood Mini Kit (Qiagene, Venlo, The Netherlands). *TPMT*2* and *TPMT*3* alleles were determined in samples with 10 ng of DNA using TaqMan SNP Genotyping Assay (Thermo Fisher Scientific, Applied Biosystems) on a Roche LightCycler 480 system, following the manufacturers' instructions. SNPs and part numbers of all TaqMan SNP Genotyping Assays are listed in Table 1.

The promoter region of *TPMT* was genotyped for VNTR in the DNA isolated from both, blood samples and LCLs. The region of interest was amplified using primers F (5′-GTCACCCGAAATCCGCCAC-3′) and R (5′-CGTTCCCTTTCTCACCCGC-3′) with the PCR reaction mix containing 80–160 ng of DNA per sample. PCR reagent from FIREPol DNA Polymerase, PCR kit, with 30% Solution S for GC rich regions (i.e., G/C rich enhancer; all from Solis BioDyne, Tartu, Estonia) was used according to the manufacturer's instructions. The following thermal cycling conditions were applied: 15 min 95°C, 30× (1 min 95°C, 30 s 58°C, 1 min 72°C), 8 min 72°C and cooling. The presence of PCR fragments containing 622–730 base pairs was confirmed by gel electrophoresis composed of 2% agarose with 1× SYBR safe DNA gel stain (Thermo Fisher Scientific, Invitrogen). The separation of fragments was performed at 100 V and 400 A and lasted 20 min. PCR products were purified by QIAquick PCR purification Kit (Qiagene, Venlo, The Netherlands) according to the manufacturer's instructions. Samples were sequenced by McLab (CA, USA).

TPMT activity assay

TPMT activity in blood samples of 29, randomly chosen, healthy individuals was measured as previously described [6].

Activity of the enzyme in LCLs was measured by the reversed-phase HPLC method [6,23]. Briefly, 10×10^6 cells were resuspended in 150 mM potassium-phosphate buffer, pH 7.4, and lysed by sonication. An aliquot of all samples was taken from the lysates for determination of proteins. The amount of produced 6MMP was measured in the rest of the cell lysate after 2-h incubation with 50 µM *S*-adenosylmethionine and 2.25 mM 6MP on 37°C.

Table 2. Frequency of *TPMT*3* genotypes in a cohort of randomly selected healthy Slovenian individuals and in the cohort enriched for *TPMT*3 alleles*.

Genotype	Randomly selected Slovenian individuals		Cohort enriched for *TPMT*3* alleles†	
	n	Frequency	n	Frequency
*3A/*3A	/	/	1	0.024
*3A/*3C	/	/	1	0.024
*1/*3A	11	0.083	32	0.268
*1/*3C	1	0.008	2	0.024
*1/*1	120	0.909	29	0.659
Total	132		65	

†Nonrandomly selected samples: lymphoblastoid cell lines from Estonian and Israeli donors (n = 41) and blood samples from Italian inflammatory bowel disease or acute lymphoblastic leukemia patients (n = 24).

Table 3. Allele frequencies of variable number of tandem repeats in the present and previous studies.

VNTR† pattern		Frequency of VNTR alleles					
		Slovenian cohort (n = 132) (present study)	Serbian ALL cohort (n = 57)	Serbian cohort ALL + normal (n = 154)	Portuguese cohort (n = 176)	Mozambique cohort (n = 88)	Asian British cohort (n = 85)
A2BC	*4a	0.439	0.535	0.542	0.447	0.506	0.482
A2B2C	*5a	0.386	0.325	0.3	0.357	0.193	0.271
A2B3C	*6a	0.087	0.053	0.075	0.096	0.034	0.147
AB4C	*6b	0.019	0	0.003	0.003	0	0
A5BC	*7a	0.015	0.053	0.036	0.056	0.102	0.029
AB5C	*7b	0.015	0	0.003	0.006	0	0
A4BC	*6c	0.008	0	0.003	0	0	0.006
A6BC	*8a	0.008	0	0.003	0.025	0.051	0.012
A2B5C	*8b	0.008	0	0	0	0.011	0
ABC	*3	0.004	0	0	0.003	0.017	0
AB2C	*4b	0.004	0.018	0.007	0.003	0.006	0.012
AB3C	*5c	0.004	0.009	0	0	0.023	0
A3B2C	*6d	0.004	0	0.016	0	0	0
A7BC	*9(a)	0.004	0	0	0	0	0.029
AB7C		0.004	0	0	0	0	0
A3BC	*5b	0	0.009	0.01	0.006	0.028	0.006
A2B4C	*7c	0	0	0	0	0.017	0
A2B6C	*9(b)	0	0	0	0	0.006	0
A4B2C	*7c	0	0	0	0	0	0.006

†Regarding the terminology of VNTR alleles, we used the nomenclature ABC instead of for example *5b, since this classification is more informative and transparent. However, both versions are specified in Table 3, allowing for comparison with previous studies.
Data was taken from the following sources: Serbian ALL cohort [18]; Serbian cohort ALL + normal [16]; Portuguese cohort [12]; Mozambique cohort [12]; Asian British cohort [26].
ALL: Acute lymphoblastic leukemia; VNTR: Variable number of tandem repeats.

Total protein concentration was determined using the BioRad Protein Assay Kit (Bio-Rad, CA, USA), in accordance with the manufacturer's instructions. The activity was calculated as the amount of 6MMP produced in an hour per mg of protein. The mean TPMT activity for each individual was obtained from at least three independent experiments.

Statistical methods

Sequences were determined using Finch TV software (Geospiza, Inc., WA, USA; http://www.geospiza.com).

LD: since SNPs and VNTRs were determined separately, we first performed an analysis to estimate the haplotypes using PHASE platform for Windows 2.1.1 [24,25], which allowed for preservation of VNTR architecture in its original form. Motifs A and B were assigned as two separate alleles in the immediate vicinity of one another (1 bp). Standardized LD (D′) was then calculated from the estimated haplotype frequencies individually for each VNTR

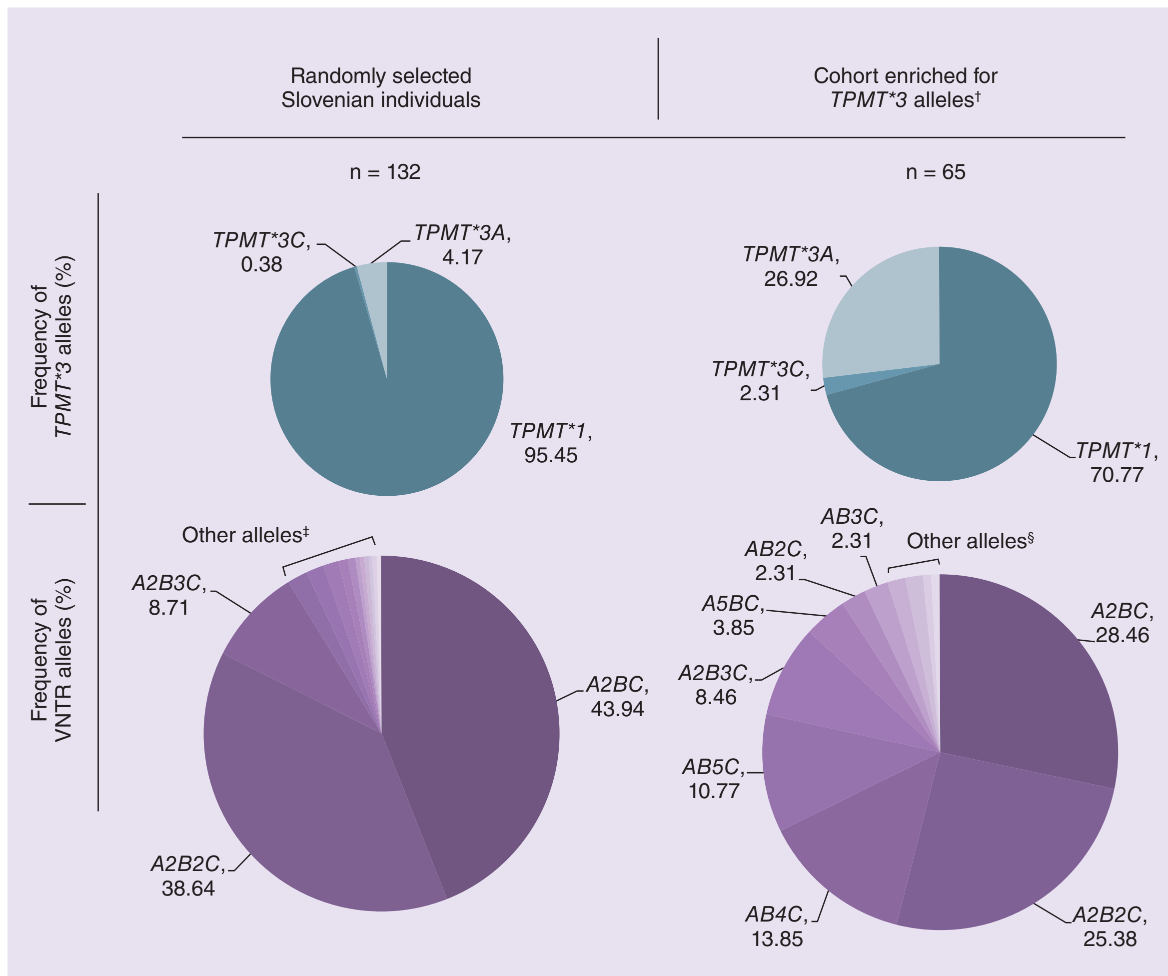


Figure 2. Allele frequency distribution diagrams for randomly and nonrandomly selected cohorts.
†Nonrandomly selected samples: lymphoblastoid cell lines from Estonian and Israeli donors (n = 41) and blood samples from Italian inflammatory bowel disease or acute lymphoblastic leukemia patients (n = 24).
‡Variable number of tandem repeats (VNTR) alleles with frequency lower than 2%: AB4C (1.89%), AB5C and A5BC (1.52%), A4BC and A6BC (0.76%), ABC, AB2C, AB3C, AB7C, A7BC and A3B2C (0.38%).
§VNTR alleles with frequency lower than 2%: A6BC and ABC (1.54%), A4BC and A2B5C (0.77%).

allele in combination with each SNP (**1, *3C, *3A* [**3B*]) (Table 4). The target VNTR allele was compared against a pool of other VNTR alleles. VNTR minisatellites have a significantly higher recombination rate than SNPs. Thus, SNPs can persist in all VNTR alleles formed in subsequent evolutionary steps. When addressing LD between VNTR patterns and *TPMT*3A* allele, we therefore merged haplotype information of samples containing VNTR alleles with ABnC, n >2 and compared them against a pool of other VNTR alleles. The significance of calculated LD was then determined by χ^2 test.

It is noteworthy that the analysis of the VNTR region and SNP genotype analyses were performed separately by two researchers in a double-blind manner, to avoid bias. Neither of the researchers knew results on the genotype of the regions prior to the conclusion of the data analysis.

Results

Frequency of *TPMT*3* & VNTR alleles

To evaluate the genetic background of TPMT in each individual, we first genotyped 132 randomly selected, healthy Slovenian blood donors for polymorphisms in coding regions and promoter of *TPMT*. The allele frequencies of *TPMT*3A* and *TPMT*3C* resembled those in other Caucasian populations. However, no *TPMT*3A* and *TPMT*3C* homozygous individuals were detected (Table 2). Furthermore, no *TPMT*2* alleles were found in the Slovenian

10.2217/pgs-2018-0123

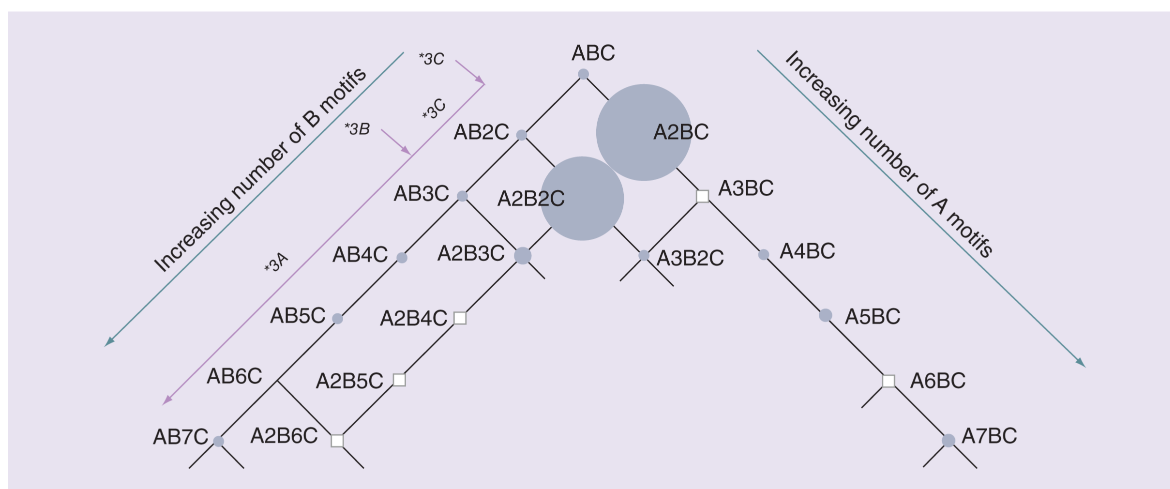


Figure 3. Phylogenetic associations between different alleles of variable number of tandem repeats in *TPMT* and their frequency distribution in the present set of Slovenian individuals, designated by the diameter of circles. □ signifies other possible variable number of tandem repeats alleles, found in lymphoblastoid cell lines or previously reported by other studies [12,16].

Table 4. Linkage disequilibrium for all combinations of variable number of tandem repeats and SNPs in *TPMT*, obtained from estimated haplotype combinations.

Allele VNTR	Allele ORF	D′	r^2	χ^2	p-value
ABC	*1	1.000	0.001	0.439	0.507
AB2C	*3C	1.000	1.000	396.000	<0.0001*
AB3C	*3A	1.000	0.078	30.726	<0.0001*
AB4C	*3A	1.000	0.469	185.725	<0.0001*
AB5C	*3A	1.000	0.362	143.417	<0.0001*
AB7C	*3A	1.000	0.019	7.623	0.006*
A2BC	*1	1.000	0.092	36.541	<0.0001*
A4BC	*1	1.000	0.001	0.442	0.506
A5BC	*1	1.000	0.003	1.345	0.246
A6BC	*1	1.000	0.001	0.590	0.442
A7BC	*1	1.000	0.000	0.146	0.702
A2B2C	*1	1.000	0.076	30.001	<0.0001*
A2B3C	*1	1.000	0.014	5.436	0.020*
A3B2C	*1	1.000	0.000	0.146	0.702
ABnC; n > 2	*3A	1.000	1.000	396.000	<0.0001*
not ABnC n > 1	*1	1.000	1.000	396.000	<0.0001*

Number of individuals: 197 (Slovenian cohort, lymphoblastoid cell lines and patients with acute lymphoblastic leukemia or inflammatory bowel disease). Significance (p) was calculated to one degree of freedom.
*Statistically significant linkage disequilibrium.
D′: Normalized linkage disequilibrium; r^2: Correlation coefficient; VNTR: Variable number of tandem repeats.

cohort. We determined 14 different VNTR alleles generating 24 different VNTR genotypes in this cohort. In accordance with previous studies, the most frequent allele was A2BC (43.1%), followed by A2B2C (37.9%) and A2B3C (8.5%) (Figure 2 & Table 3). All other alleles were represented at frequencies of less than 2% in the population. The most common genotypes were, as expected, A2BC/A2B2C (37.3%), A2BC/A2BC (17.9%) and A2B2C/A2B2C (14.2%). Interestingly, the VNTR allele AB2C was present only in the individual with allele *TPMT*3C*. Moreover, all *TPMT*3A* individuals carried one of the following VNTR alleles: AB3C, AB4C, AB5C and AB7C. None of these VNTR alleles (ABnC, n ≥2) were present in wild-type individuals. To our knowledge, the VNTR allele AB7C was observed for the first time.

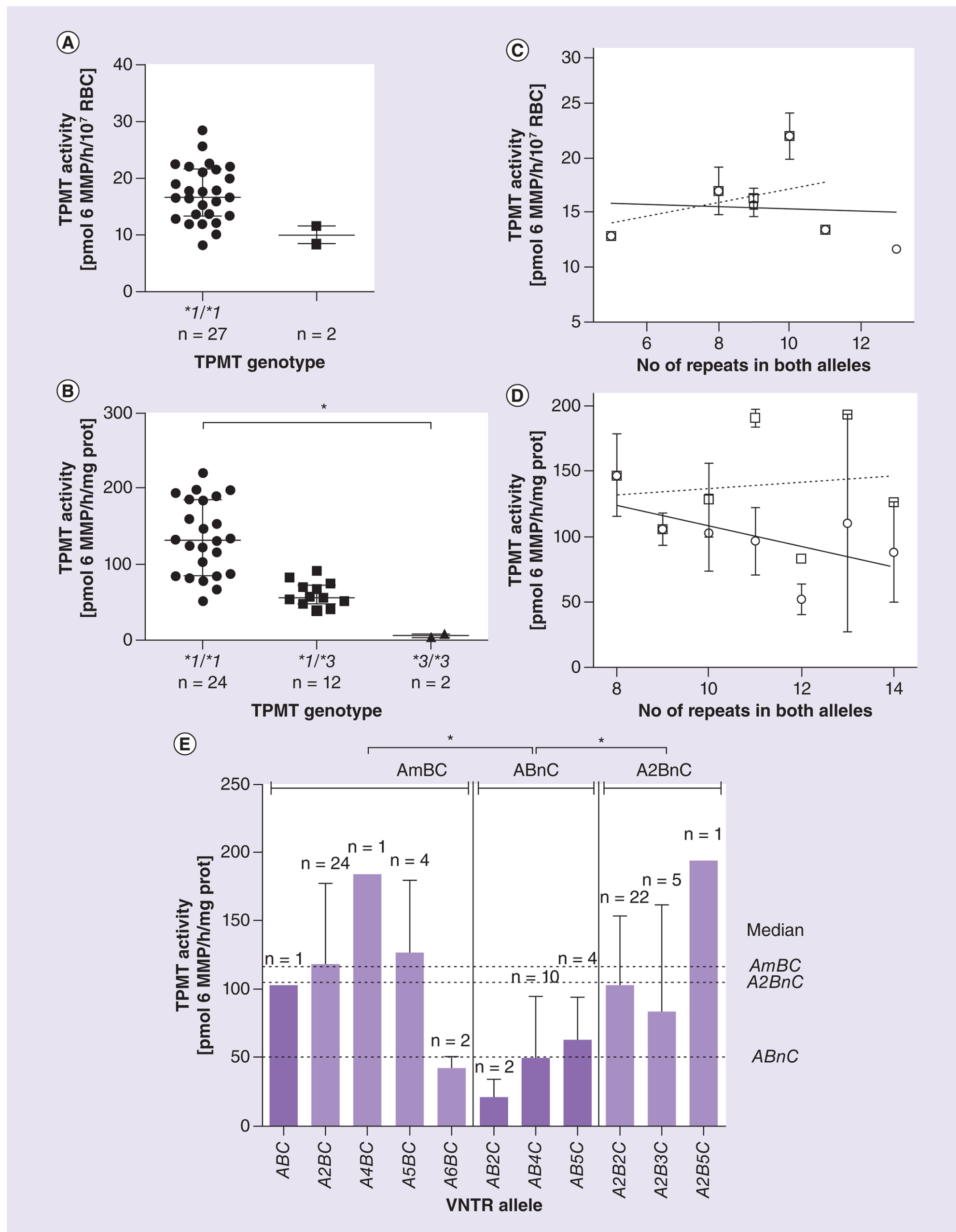


Figure 4. Activity of TPMT in relation to *TPMT*3* and alleles of variable number of tandem repeats. (A) Distribution of TPMT activity according to *TPMT*3* genotype in Slovenian individuals. **(B)** Classification of TPMT activity based on TPMT genotype in lymphoblastoid cell lines (LCLs). Groups of LCLs with different *TPMT*3* genotype significantly differ in TPMT activity (p = 0.0001). **(C)** and **(D)** TPMT activity of healthy Slovenian individuals and LCLs, respectively, in relation to the number of repeats in the VNTR pattern. TPMT activity trend with the total number of repeats is represented by ○, solid line, while the trend of only wild-type (**1*/**1*) individuals and LCLs is shown by □, dashed line. **(E)** Distribution of TPMT activity of LCLs across the groups with different VNTR patterns. LCLs carrying at least one *TPMT*3* allele are shown in darker purple. Pooled subgroups of variable number of tandem repeats motifs (i.e., AmBC, ABnC and A2BnC) exhibited statistically significant differences in TPMT activity (p = 0.0001). Results are presented as median ± interquartile difference.
*p = 0.0001.

10.2217/pgs-2018-0123

As the frequency of *TPMT*3* alleles in the random population is too low to convincingly confirm the association between *TPMT*3* and VNTR alleles, we next selected additional samples from individuals previously determined to carry *TPMT*3* allele(s). Additionally, a few samples of wild-type *TPMT* individuals were added. These samples originated from IBD or ALL patients (n = 24) and from LCLs obtained from Estonian and Israeli donors (n = 41). By combining samples from LCLs, IBD and ALL patients, we intentionally created a cohort with elevated frequency of *TPMT*3* alleles (Table 2 & Figure 2). DNA was extracted from 2 *TPMT*1/*3C* heterozygous, 32 *TPMT*1/*3A* heterozygous, one homozygous and one mixed heterozygous (**3A/*3A* and **3A/*3C*), as well as 29 wild-type individuals, and *TPMT*3* genotype was reconfirmed (Table 2). The sequencing of the *TPMT* promoter region determined 12 structurally different VNTR alleles in this *TPMT*3* enriched cohort (Figure 2).

In all three cohorts combined (healthy Slovenian, Italian ALL and IBD patients and LCLs), AB3C, AB4C, AB5C and AB7C patterns were detected only in individuals with at least one *TPMT*3A* allele. Furthermore, the individual with genotype **3A/*3A* was also AB4C/AB4C homozygous at the VNTR locus. These results suggest that *TPMT*3A* allele is in linkage with ABnC (n $\geq$3) VNTR patterns and corroborate results from Slovenian population study. In all three cohorts, AB2C was detected in individuals with **1/*3C* and **3A/*3C* genotypes. The latter had AB4C/AB2C genotype at the VNTR locus. Since AB4C associates with the *TPMT*3A* allele, the data further suggest that the *TPMT*3C* allele corresponds to AB2C VNTR motif. The high percentage of ABnC alleles in the cohort enriched for *TPMT*3* variants (Figure 2) confirmed the association between VNTR motifs AB2C and ABnC (n $\geq$3) and *TPMT*3C* and *TPMT*3A* alleles, respectively.

Evolutionary linkage of VNTR & *TPMT*3* polymorphisms

To statistically evaluate the correlation between *TPMT*3* and VNTR alleles, we performed an analysis of LD, which further confirmed our hypothesis. Genotype data of merged samples from the Slovenian cohort, LCLs, ALL and IBD patients were used to estimate the haplotypes, and calculate the LD and its level of significance (Table 4). VNTR polymorphism AB2C was in statistically significant complete LD with *TPMT*3C* polymorphism ($D' = 1$, $r^2 = 1$; $p < 0.0001$), and pooled polymorphisms AB3C, AB4C, AB5C and AB7C were in complete LD with *TPMT*3A* polymorphism ($D' = 1$, $r^2 = 1$; $p < 0.0001$). Other VNTR patterns were in statistically significant LD with wild-type (*TPMT*1*) allele (Table 4).

All *TPMT*3* alleles occur with VNTR pattern ABnC, where n is at least two. Therefore, we investigated the evolutionary basis for such an association. *TPMT*3A* allele is a haplotype that contains *TPMT *3C* and *TPMT *3B* SNPs in a coding region of *TPMT* (Figure 1) [27]. *TPMT*3C* allele is quite common, especially in populations of African origin [28,29]. Meanwhile, *TPMT *3B* allele almost never occurs separately from *TPMT*3C* [30,31]. In all analyzed samples, *TPMT *3C* allele was present together with VNTR polymorphisms ABnC with two B motifs (n = 2) and *TPMT*3A* allele co-existed with the same VNTR patterns that contained three or more B motifs (n >2). Assuming a stepwise mutation model as an evolutionary description of VNTR formation and their phylogenetic evaluation [12], we hypothesize that *TPMT*3C* variation occurred concurrently with the AB2C polymorphism (Figure 3). In further evolutionary genetic changes, an additional B motif in the VNTR region was gained. At that time, the *TPMT*3B* mutation appeared, creating the *TPMT*3A* haplotype, suggesting that the *TPMT*3B* variation occurred sequentially after *TPMT*3C* and not in parallel or separately from it. Since the mutation rate of VNTR polymorphisms is significantly higher compared with the mutation rate of variations resulting in SNPs, *TPMT*3A* was retained, while the VNTR region gained additional B motifs, generating several haplotypes with *TPMT*3A* and VNTR ABnC polymorphisms (Figure 3).

Influence of *TPMT*3* & VNTR alleles on the TPMT activity

To evaluate whether VNTR polymorphisms affect TPMT activity, we measured the basal activity of the enzyme in 29 of 132 healthy Slovenian volunteers and in 38 of 41 LCLs. Median TPMT activity of Slovenian donors with wild-type *TPMT* genotype (**1/*1*) was 16.6 pmol 6MMP/h/10^7 RBC, whereas the median activity of heterozygous (**1/*3*) individuals was 10.0 pmol 6MMP/h/10^7 RBC (Figure 4A). Due to the low frequency of *TPMT*3* alleles in the randomly selected population, the size of the Slovenian cohort was too small to allow the statistical analysis of TPMT activity distribution (Figure 4A). However, the activity of two heterozygous individuals was in the range of previously reported intermediate TPMT activities in the Slovenian and other populations [6,32,33]. The broad distribution of activities in individuals with wild-type *TPMT* is also in accordance with several previously published studies [1,34]. TPMT activity, measured in LCL lysates, significantly correlated with the number of *TPMT*3* alleles ($p < 0.0001$; Figure 4B). LCLs with wild-type *TPMT* (**1/*1*) had the highest activity (median

137.4 pmol 6MMP/h/mg prot.), heterozygous LCLs (**1/*3*) an intermediate level of activity (median 56.2 pmol 6MMP/h/mg prot.) and variant homozygotes (**3/*3*) had the lowest TPMT activity (5.73 pmol 6MMP/h/mg prot.).

In accordance with previous studies [9,14], we observed a decrease in median TPMT activity with an increase in the number of repeats in both alleles, irrespective of the number of A and B motifs (Figure 4C & D, solid line). The same trend of decreasing TPMT activity with the number of VNTR motifs was absent in only wild-type (**1/*1*) individuals and LCLs (Figure 4C & D, dashed line). In addition, we noticed a trend of fewer LCLs with **1/*1* genotype within the groups with more than ten tandem repeats, suggesting that individuals with variant *TPMT*3* allele also have longer VNTR region. Not only the total number of tandem repeats, but also their VNTR patterns (i.e., number of A and B motifs) were associated with TPMT activity (Figure 4E). ABnC patterns had significantly lower enzyme activity compared with A2BnC ($p < 0.0001$) and AmBC ($p < 0.0001$) patterns (Figure 4E).

Discussion

The present study finds that there is a LD between VNTR patterns with ABnC structure ($n \geq 2$) and *TPMT*3* alleles. Specifically, motif AB2C is associated with *TPMT*3C* and ABnC ($n \geq 3$) motifs with *TPMT*3A* allele. Furthermore, we identified a previously unreported VNTR pattern AB7C. The individual with this pattern is a carrier of *TPMT*3A* allele. Further functional studies for this individual (e.g., the analysis of TPMT activity) could not have been performed, as we have not been able to obtain the red blood cell sample from this individual.

Our data are strongly supported by other studies. The most demonstrable linkage disequilibria observed in previous studies were between *TPMT *3A* and AB4C (*6b), and between *TPMT *3C* and AB2C (*4b) [26]. In that study all individual subjects (who originated from Pakistan and India) had only the AB2C VNTR pattern: no other ABnC polymorphisms were detected [26]. This is in accordance with the distribution of *TPMT*3* alleles in the ethnic groups of Asia [35]. In another study, which assessed a Mozambican population, VNTR motifs AB4C (*V6b) and AB5C (*V7b) were absent, while AB2C motif was present; in this population *TPMT*3C* is the prevalent allele and *TPMT*3A* is rare [12]. This is in line with our results and in accordance with the distribution of these variants in populations of African and Asian origin [12]. Furthermore, *VNTR*6*, which could represent either AB4C, A4BC or A2B3C motifs, has been linked to the *TPMT *3A* polymorphism in early studies of VNTR alleles in *TPMT* [9,12]. Unfortunately, the information on qualitative structure of VNTR polymorphisms was lacking at that time, making it impossible to determine which of the three VNTR patterns was in linkage with *TPMT*3A* allele. Based on these findings, we can predict that African population would have extremely low frequencies of ABnC, $n \geq 3$, alleles but higher frequency of AB2C allele, compared with the Caucasian population. Populations of Asian origin have a low frequency of *TPMT*3C* and an even lower frequency of *TPMT*3A* [35,36]. Therefore, AB2C with similar frequency to the Caucasian population and low frequency of ABnC, $n \geq 3$, would be expected in these populations.

To compare our results with those in 1000 Genomes Project, we endeavored to analyze the data from the database [37], using NCBI Genome Browser for GRCh37.p13 [38]. The data confirmed population genetics and frequency distribution of *TPMT*3C* and *TPMT*3A* alleles, and complete LD between *TPMT*3C* and *TPMT*3B* alleles. Individuals with *TPMT*3* alleles were further analyzed for the VNTR pattern. The coverage of this region, surrounding nucleotide 18,155,419 (Chr 6), was very low or nonexistent, which in the majority of cases made it impossible to identify these long and complex VNTR polymorphisms lying within a G/C-rich region in the *TPMT* promoter. Nevertheless, in approximately 20% of *TPMT*3* individuals the VNTR pattern in the *TPMT* promoter could be partially predicted with the sequence pointing toward AB- and not AA- pattern, which is in line with our study, where patterns with single A and several B VNTR motifs occur in individuals carrying *TPMT*3* allele. . Despite the great advances of the next-generation sequencing, it has been previously reported, that genome assembly for highly G/C-rich regions is challenging and results in reduced or no coverage [39]. Alternative approaches in sequencing complete VNTR assemblies such as single molecule real time sequencing have the potential to facilitate future high throughput studies of such polymorphisms [40].

Conclusion

We report a novel VNTR pattern AB7C, which was observed in a *TPMT*1/*3A* individual. Furthermore, our findings provide novel insights into the concurrent development of VNTR and *TPMT*3* alleles. We confirmed that *TPMT*3* alleles correlate with TPMT activity. However, we could not conclude what influence (if any) VNTR polymorphisms have on the activity of the enzyme, since patterns with ABnC, $n \geq 2$, are in significant LD with *TPMT*3* alleles. Our study suggests that VNTR pattern AB2C is associated with *TPMT*3C*, while patterns

ABnC, n ≥3, are associated with *TPMT*3A* allele, suggesting these VNTR sequences as indirect pharmacogenomic markers.

Future perspective

Based on the present results, it is possible that the VNTR region has no influence on the TPMT activity and that the previously observed association is the sole consequence of the linkage between ABnC, n ≥2, VNTR motifs and *TPMT*3* alleles. Conversely, the same results could indicate the joint influence of the VNTR region and *TPMT*3* alleles on the enzyme activity. Interesting insights in function of *TPMT* promoter region have been previously reported in terms of TPMT expression and binding of transcription factors [10,16,41]; however, the role of VNTR in the promoter region of *TPMT* remains unresolved. Further functional *in vitro* studies are necessary to decipher the possible influence of the VNTR region on TPMT activity.

Based on its sequence, the B motif is strongly indicated as predicted binding site for transcription factor Sp1, which is a recognized regulator of *TPMT* transcription [16,41]. At the same time, sequence of motif B carries determinants that enable formation of G-quadruplexes [42]. Even though their formation in promoter of *TPMT* has not been investigated, G-quadruplexes can be envisaged as regulators of VNTR stability or an element that directs transcription of the gene. In light of our findings, future studies of such secondary structures would be extremely important, since multiple B motifs are present in the promoter regions of *TPMT*3* alleles. They might be even assumed as regulatory domains counteracting TPMT-activity-lowering SNPs in coding region of the *TPMT*.

Summary points

- We report a novel pattern of variable number of tandem repeats (VNTR) in the *TPMT* promoter: AB7C.
- Only *TPMT*3A* individuals carried one of the following VNTR alleles: AB3C, AB4C, AB5C and AB7C.
- Only *TPMT*3C* individuals carried AB2C allele.
- VNTR patterns with single A and multiple B motifs are associated with *TPMT*3* alleles. VNTR patterns with three or more B motifs developed concurrently with the *TPMT*3B* SNP from pattern AB2C, which coexisted with the *TPMT*3C* allele.

Author's contributions

D Urbančič, I Mlinarič-Raščan and N Karas Kuželički designed research. D Urbančič and A Šmid performed research. D Urbančič analyzed data. I Mlinarič-Raščan, G Stocco and G Decorti contributed reagents/clinical samples/analysis tools. D Urbančič, A Šmid and N Karas Kuželički wrote the paper.

Acknowledgements

The authors wish to thank all participants who entered the study cohorts and clinical coordinators, M Rabusin and M Bramuzzo from the Institute for Maternal and Child Health IRCCS and BG Trieste for overseeing the recruitment of ALL and IBD patients. We would like to thank D Gurwitz from the Department of Human Molecular Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv and A Metspalu from The Estonian Genome Center, University of Tartu for providing *in vitro* LCL model. We thank R Franca from the Institute for Maternal and Child Health IRCCS Burlo Garofolo Trieste for technical assistance and coordinating ALL repository, and M Lucafò from the Department of Medical, Surgical and Health Sciences, University of Trieste for coordinating IBD repository. The authors thank P Ferkov from the Faculty of Pharmacy, University of Ljubljana, D Selvestrel and O Montecchini from the Developmental and Reproductive Sciences, University of Trieste for technical assistance, and M Milek for providing useful information on TPMT in Slovene population. We would like to thank TJ Bevec for the proof reading of this manuscript.

Financial & competing interests disclosure

This work was supported by the Slovenian Research Agency (J3-6792, P1-0208, MR34512). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

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

Ethical disclosure

The authors state that, 'the study has been approved by Local and National Medical Ethics Committees and has followed the principles outlined in the Declaration of Helsinki for all human experimental investigations. Written informed consent has been obtained from all participating subjects. The experiments comply with the current laws of the Republic of Slovenia and the Italian Republic'.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Lennard L, Cartwright CS, Wade R, Vora A. Thiopurine methyltransferase and treatment outcome in the UK acute lymphoblastic leukaemia trial ALL2003. *Br. J. Haematol.* 170(4), 550–558 (2015).
2. Weinshilboum RM, Sladek SL. Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity. *Am. J. Hum. Genet.* 32(5), 651–662 (1980).
- • **The first study describing genetic background of thiopurine *S*-methyltransferase (TPMT) activity.**
3. Lennard L, Cartwright CS, Wade R, Vora A. Thiopurine dose intensity and treatment outcome in childhood lymphoblastic leukaemia: the influence of thiopurine methyltransferase pharmacogenetics. *Br. J. Haematol.* 169(2), 228–240 (2015).
- •• **One of the recent studies sill confirming importance of implementation of TPMT determination in clinical practice.**
4. Lee M-N, Kang B, Choi SY *et al.* Relationship between azathioprine dosage, 6-thioguanine nucleotide levels, and therapeutic response in pediatric patients with IBD treated with azathioprine. *Inflamm. Bowel Dis.* 21(5), 1054–1062 (2015).
5. Liang JJ, Geske JR, Boilson BA *et al.* TPMT genetic variants are associated with increased rejection with azathioprine use in heart transplantation. *Pharmacogenet. Genomics* 23(12), 658–665 (2013).
6. Milek M, Murn J, Jaksic Z, Lukac Bajalo J, Jazbec J, Mlinaric Rascan I. Thiopurine *S*-methyltransferase pharmacogenetics: genotype to phenotype correlation in the Slovenian population. *Pharmacology* 77(3), 105–114 (2006).
7. Tamm R, Mägi R, Tremmel R *et al.* Polymorphic variation in TPMT is the principal determinant of TPMT phenotype: a meta-analysis of three genome-wide association studies. *Clin. Pharmacol. Ther.* 101(5), 684–695 (2017).
8. Lee M-N, Kang B, Choi SY *et al.* Impact of genetic polymorphisms on 6-thioguanine nucleotide levels and toxicity in pediatric patients with IBD treated with azathioprine. *Inflamm. Bowel Dis.* 21(12), 2897–2908 (2015).
9. Spire-Vayron de la Moureyre C, Debuysere H, Mastain B *et al.* Genotypic and phenotypic analysis of the polymorphic thiopurine *S*-methyltransferase gene (TPMT) in a European population. *Br. J. Pharmacol.* 125(4), 879–887 (1998).
10. Spire-Vayron de la Moureyre C, Debuysère H, Fazio F *et al.* Characterization of a variable number tandem repeat region in the thiopurine *S*-methyltransferase gene promoter. *Pharmacogenetics* 9(2), 189–198 (1999).
11. Vergnaud G, Denoeud F. Minisatellites: mutability and genome architecture. *Genome Res.* 10(7), 899–907 (2000).
12. Alves S, Amorim A, Prata MJ. Evolution of a VNTR located within the promoter region of the thiopurine methyltransferase gene: inferences from population and sequence data. *Hum. Genet.* 111(2), 172–178 (2002).
- •• **Proposes single-step mutation model for polymorphisms of variable number of tandem repeats (VNTR) in promoter region of *TPMT*.**
13. Kimura M, Ohta T. Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc. Natl Acad. Sci. USA* 75(6), 2868–2872 (1978).
14. Yan L, Zhang S, Eiff B *et al.* Thiopurine methyltransferase polymorphic tandem repeat: genotype–phenotype correlation analysis. *Clin. Pharmacol. Ther.* 68(2), 210–219 (2000).
15. Alves S, Amorim A, Ferreira F, Prata MJ. Influence of the variable number of tandem repeats located in the promoter region of the thiopurine methyltransferase gene on enzymatic activity. *Clin. Pharmacol. Ther.* 70(2), 165–174 (2001).
16. Zukic B, Radmilovic M, Stojiljkovic M *et al.* Functional analysis of the role of the *TPMT* gene promoter *VNTR* polymorphism in *TPMT* gene transcription. *Pharmacogenomics* 11(4), 547–557 (2010).
- • **The scientific paper that underlines the importance of thoroughly annotated VNTR architecture.**
17. Kotur N, Stankovic B, Kassela K *et al.* 6-mercaptopurine influences *TPMT* gene transcription in a *TPMT* gene promoter variable number of tandem repeats-dependent manner. *Pharmacogenomics* 13(3), 283–295 (2012).
18. Kotur N, Dokmanovic L, Janic D *et al.* *TPMT* gene expression is increased during maintenance therapy in childhood acute lymphoblastic leukemia patients in a *TPMT* gene promoter variable number of tandem repeat-dependent manner. *Pharmacogenomics* 16(15), 1701–1712 (2015).
- •• **The most important recent advances in functional and clinical characterization of the promoter region of *TPMT*, suggesting VNTR alleles influence expression of *TPMT* in patients treated with 6-mercaptopurine.**
19. Stocco G, Martelossi S, Arrigo S *et al.* Multicentric case-control study on azathioprine dose and pharmacokinetics in early-onset pediatric inflammatory bowel disease. *Inflamm. Bowel Dis.* 23(4), 628–634 (2017).
20. Stocco G, Martelossi S, Barabino A *et al.* Glutathione-*S*-transferase genotypes and the adverse effects of azathioprine in young patients with inflammatory bowel disease. *Inflamm. Bowel Dis.* 13(1), 57–64 (2007).
21. Morag A, Kirchheiner J, Rehavi M, Gurwitz D. Human lymphoblastoid cell line panels: novel tools for assessing shared drug pathways. *Pharmacogenomics* 11(3), 327–340 (2010).
22. Leitsalu L, Haller T, Esko T *et al.* Cohort profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int. J. Epidemiol.* 44(4), 1137–1147 (2015).

23. Keizer-Garritsen JJ, Brouwer C, Lambooy LHJ *et al.* Measurement of thiopurine *S*-methyltransferase activity in human blood samples based on high-performance liquid chromatography: reference values in erythrocytes from children. *Ann. Clin. Biochem.* 40(Pt 1), 86–93 (2003).

24. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68(4), 978–989 (2001).

25. Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am. J. Hum. Genet.* 76(3), 449–462 (2005).

26. Marinaki AM, Arenas M, Khan ZH *et al.* Genetic determinants of the thiopurine methyltransferase intermediate activity phenotype in British–Asians and Caucasians. *Pharmacogenetics* 13(2), 97–105 (2003).

27. Szumlanski C, Otterness D, Her C *et al.* Thiopurine methyltransferase pharmacogenetics: human gene cloning and characterization of a common polymorphism. *DNA Cell Biol.* 15(1), 17–30 (1996).

28. Hon YY, Fessing MY, Pui CH, Relling MV, Krynetski EY, Evans WE. Polymorphism of the thiopurine *S*-methyltransferase gene in African–Americans. *Hum. Mol. Genet.* 8(2), 371–376 (1999).

29. Hiratsuka M, Inoue T, Omori F, Agatsuma Y, Mizugaki M. Genetic analysis of thiopurine methyltransferase polymorphism in a Japanese population. *Mutat. Res.* 448(1), 91–95 (2000).

30. Rossino R, Vincis C, Alves S *et al.* Frequency of the thiopurine *S*-methyltransferase alleles in the ancient genetic population isolate of Sardinia. *J. Clin. Pharm. Ther.* 31(3), 283–287 (2006).

31. Bahari A, Hashemi M, Bari Z, Moazeni-Roodi A, Kaykhaei M-A, Narouie B. Frequency of thiopurine *S*-methyltransferase (TPMT) alleles in southeast Iranian population. *Nucleosides Nucleotides Nucleic Acids* 29(3), 237–244 (2010).

32. Karas-Kuželički N, Šmid A, Tamm R, Metspalu A, Mlinarič-Raščan I. From pharmacogenetics to pharmacometabolomics: SAM modulates TPMT activity. *Pharmacogenomics* 15(11), 1437–1449 (2014).

33. Chouchana L, Narjoz C, Roche D *et al.* Interindividual variability in TPMT enzyme activity: 10 years of experience with thiopurine pharmacogenetics and therapeutic drug monitoring. *Pharmacogenomics* 15(6), 745–757 (2014).

34. Hindorf U, Appell ML. Genotyping should be considered the primary choice for pre-treatment evaluation of thiopurine methyltransferase function. *J. Crohns Colitis* 6(6), 655–659 (2012).

35. Kubota T, Chiba K. Frequencies of thiopurine *S*-methyltransferase mutant alleles (*TPMT*2*, **3A*, **3B* and **3C*) in 151 healthy Japanese subjects and the inheritance of *TPMT*3C* in the family of a propositus. *Br. J. Clin. Pharmacol.* 51(5), 475–477 (2001).

36. Lu Y, Kham SK-Y, Foo T-C *et al.* Genotyping of eight polymorphic genes encoding drug-metabolizing enzymes and transporters using a customized oligonucleotide array. *Anal. Biochem.* 360(1), 105–113 (2007).

37. 1000 Genomes Project Consortium; Auton A, Brooks LD *et al.* A global reference for human genetic variation. *Nature* 526(7571), 68–74 (2015).

38. Geer LY, Marchler-Bauer A, Geer RC *et al.* The NCBI BioSystems database. *Nucleic Acids Res.* 38(Database issue), D492–D496 (2010).

39. Chen Y-C, Liu T, Yu C-H, Chiang T-Y, Hwang C-C. Effects of GC bias in next-generation-sequencing data on *de novo* genome assembly. *PLoS ONE* 8(4), e62856 (2013).

40. Wenzel A, Altmueller J, Ekici AB *et al.* Single molecule real time sequencing in ADTKD- MUC1 allows complete assembly of the VNTR and exact positioning of causative mutations. *Sci. Rep.* 8(1), 4170 (2018).

41. Fessing MY, Krynetski EY, Zambetti GP, Evans WE. Functional characterization of the human thiopurine *S*-methyltransferase (TPMT) gene promoter. *Eur. J. Biochem.* 256(3), 510–517 (1998).

42. Todd AK, Johnston M, Neidle S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* 33(9), 2901–2907 (2005).