Cancer Chemotherapy and Pharmacology (2019) 83:875–880
https://doi.org/10.1007/s00280-019-03803-8

ORIGINAL ARTICLE

Classification and regression tree-based prediction of 6-mercaptopurine-induced leucopenia grades in children with acute lymphoblastic leukemia

Shaik Mohammad Naushad[1] · Patchava Dorababu[2] · Yedluri Rupasree[1] · Addepalli Pavani[1] · Digumarti Raghunadharao[3] · Tajamul Hussain[4,5] · Salman A. Alrokayan[5,6] · Vijay Kumar Kutala[7]

Received: 22 October 2018 / Accepted: 22 February 2019 / Published online: 26 February 2019

Abstract

Purpose The rationale of the current study was to develop 6-mercaptopurine (6-MP)-mediated hematological toxicity prediction model for acute lymphoblastic leukemia (ALL) therapeutic management.

Methods A total of 96 children with ALL undergoing therapy with MCP-841 protocol were screened for all the ten exons of TPMT, exon 2, exon 3 and intron 2 of ITPA using bidirectional sequencing. This dataset was used to construct prediction models of leucopenia grade by constructing classification and regression trees (CART) followed by smart pruning.

Results The developed CART model indicated TPMT*12 and TPMT*3C as the key determinants of toxicity. TPMT int3, int4 and int7 polymorphisms exert toxicity when co-segregated with one mutated allele of TPMT*12 or TPMT*3C or ITPA exon 3. The developed CART model exhibited 93.6% accuracy in predicting the toxicity. The area under the receiver operating characteristic curve was 0.9649.

Conclusions TPMT *3C and TPMT*12 are the key determinants of 6-MP-mediated hematological toxicity while other variants of TPMT (int3, int4 and int7) and ITPA ex2 interact synergistically with TPMT*3C or TPMT*12 variant alleles to enhance the toxicity. TPMT and ITPA variants cumulatively are excellent predictors of 6-MP-mediated toxicity.

Keywords Acute lymphoblastic leukemia · 6-Mercaptopurine · Thiopurine methyl transferase · Glutamate carboxypeptidase II · Machine learning · Leucopenia grade

Introduction

Purine analogue, 6-mercaptopurine (6-MP), is part of the poly chemotherapeutic regimen used for the maintenance therapy of acute lymphoblastic leukemia (ALL). In combination with methotrexate, it is used to maintain clinical remission for 2–3 years. The tailoring of 6-MP dose is based on total leukocyte count. Non-enzymatic removal of nitroimidazole group from azathioprine results in the formation of 6-MP, which undergoes a series of reactions to form thioinosine monophosphate (TIMP), thioxanthine monophosphate (TXMP), thioguanosine monophosphate (TGMP) and 6-thioguanine (6-TG) in the presence of hypoxanthine phosphoribosyltransferase 1 (HPRT),

Shaik Mohammad Naushad and Patchava Dorababu equal first authorship.

✉ Shaik Mohammad Naushad
naushadsm@gmail.com

[1] Head-Biochemical Genetics and Pharmacogenomics, Sandor Speciality Diagnostics Pvt Ltd, Banjara Hills, Road No 3, Hyderabad 500034, India

[2] Department of Pharmacology, Apollo Institute of Medical Sciences and Research, Hyderabad, India

[3] Homibhabha Cancer Hospital and Research Centre, Aganampudi, Visakhapatnam, India

[4] Center of Excellence in Biotechnology Research, College of Science, King Saud University, Riyadh, Saudi Arabia

[5] Research Chair for Biomedical Applications of Nanomaterials, Department of Biochemistry, College of Science, King Saud University, Riyadh, Saudi Arabia

[6] Biochemistry Department, College of Science, King Saud University, Riyadh, Saudi Arabia

[7] Department of Clinical Pharmacology and Therapeutics, Nizam's Institute of Medical Sciences, Hyderabad, India

inosine-5′-monophosphate dehydrogenase (IMPDH), guanine monophosphate synthase (GMPS) enzymes, respectively. Thiopurine methyltransferase (TPMT) catalyzes methylation of 6-MP, TIMP, TGMP and 6-TG to form inactive drug metabolites and hence confers protection against 6-MP drug toxicity. It requires S-adenosylmethionine as the co-substrate for these methylation reactions.

Higher median DNA-thioguanine levels were reported in TPMT and inosine triphosphate pyrophosphatase (ITPA) ITPA low-activity patients when compared to wild-type patients, which in turn positively correlated with median erythrocyte thioguanine levels [1]. Patients with TPMT mutations were found to be at higher risk of hematological toxicity requiring more 6-MP dose adjustments in homozygous mutants compared to heterozygous [2]. Although TPMT activity was shown to increase by 20% following thiopurine therapy, even TPMT heterozygous subjects exhibited lower TPMT activity than the subjects with TPMT wild genotype thus contributing to hematological toxicity [3]. TPMT*2, *3B and *3C variants were reported to exhibit higher 6-TGN and lower 6-methyl MP levels compared to wild type [4]. Earlier, we have reported the association of thiopurine [5] and folate pathway [6] genetic variants with 6-mercaptopurine-induced hematological toxicity in children with ALL. Exogenous S-adenosyl methionine was reported to rescue the cells from the toxic effects of 6-MP by delaying the apoptosis [7].

The rationale of the current study was to construct a classification and regression tree (CART)-based machine learning algorithm for prediction of leucopenia grades using our earlier data on 6-mercaptopurine-induced hematological toxicity. Such model may be helpful for the medical oncologist in quick decision making.

Materials and methods

Recruitment of patients

A total of 96 children (70 boys and 26 girls) with acute lymphoblastic leukemia undergoing treatment as per the MCP-841 protocol were recruited for the study at Department of Medical Oncology, Nizam's Institute of Medical Sciences, Hyderabad, India. The mean age of the patients was 9.8 ± 5.0 year. All the patients were followed up for the treatment outcome for a period of 2 years. The study protocol was approved by the Institutional Ethical Committee of Nizam's Institute of Medical Sciences (NIMS), Hyderabad, India (EC/NIMS/1123/2009). Informed consent was obtained from the parents or guardian of each subject.

Therapeutic intervention

During the maintenance phase of the treatment, an oral dose of 75 mg/m^2 of 6-MP was given daily for 3 weeks, skipping every fourth week, for a total duration of 12 weeks. Six such maintenance cycles were given to each child. A complete hemogram was performed on day 43, day 71, and day 99 of the maintenance phase and the average decrease in total leukocyte count (TLC) during these three occasions was taken as a measure of toxicity. If any patient had drug-related toxicity, then the dose of 6-MP was reduced by 10–20%. The toxicity grading was done as per the Common Toxicity Criteria (CTC) version 3 recommended by National Cancer Institute, USA for classifying hematological toxicity. These grades explain the severity of toxicity as depicted below: Grades 1—Mild, Grade 2—Moderate, Grade 3—Severe, Grade 4—Life-threatening, and Grade 5—Death. During the 6-MP therapy, a dose of methotrexate (MTX), 15 mg/m^2 PO, was given once a week for 3 weeks. The doses of both drugs were "titrated" to keep the white blood cell (WBC) count in the range of 2000–3000 Cu/mm3.

At the time of recruitment, demographic details and medical history of all the patients were recorded. The demographic details include age, gender, and body surface area. The medical history includes types of ALL (T-ALL or B-ALL), relapse or death etc., were obtained.

Genetic analysis

Whole blood samples were collected in EDTA and Genomic DNA was extracted using the standard phenol–chloroform extraction method after proteinase K digestion. All the ten exons of TPMT gene; exon 2, exon 3, and intron 2 of ITPA gene were amplified using gene-specific primers in 49 subjects. As exon 2, exon 3 and intron 2 of ITPA harbor majority of the reported variants of ITPA, only these regions were targeted for the analysis. Primers were designed in-house and the obtained amplicons were subjected to bidirectional sequencing using Big Dye Taq FS terminator method on ABI (Applied Biosystems, USA) 96-capillary 3730XL sequencer [5]. While the data were available from the previous studies [5, 6].

Development of machine learning algorithm

All the genotype data were computed in 0, 1 and 2 format based on the number of variant alleles thus represents wild, heterozygous and homozygous mutant genotypes. Age, gender, genetic variants were used as input variables while leucopenia grade was used as the output variable. TPMT*3C, TYMT*12 (rs2842934), TPMT 12356 C>T (rs4449636),

TPMT 16606 T>C (rs2518463), TPMT 26354 G>T (rs2842949), ITPA 94 C>A (rs1127354), ITPA 138 G>A (rs8362) were observed in this cohort.

We have employed machine learning based construction of classification and regression tree (CART) followed by smart pruning for the development of the model (http://www.bigml.com). Only 75% of the data were used to train the model and rest of 25% of the data were used for testing and validation. Briefly, the digitized data with input and output variables were computed. We have used hematological toxicity grade as the 'response variable' that has four classes 1, 2, 3 and 4. The most significant contributor to toxicity grade was identified as the 'Root node' and splitting of the root node was performed based on homogeneity using a purity criterion that can differentiate the response variable much more precisely. The input variables, i.e., different genotypes contributing towards the response variable, i.e., toxicity grade were added to sequentially to the growing tree in descending order of importance. This recursive partitioning tree had a cell of partition at each node. Final split or node was termed as a 'leaf'. Through smart pruning, we have minimized the complexity of the model and avoided overfitting of the data. Since the output was leucopenia grade in terms of 1, 2, 3 and 4; true assessment of lower grades, i.e., 1 and 2 was considered true negative (TN), while true assessment of higher grades, i.e., 3 and 4 was considered true positive (TP). Predicting a lower grade as higher was considered false positive (FP) while predicting a higher grade as lower was considered false negative (FN). TN, TP, FP and FN values were computed in 2×2 contingency table to calculate overall accuracy, and receiver operating characteristic curve was plotted to assess its clinical utility.

Statistical analysis

Univariate analysis was performed using logistic regression analysis where in output is computed in 0 (toxicity grade 1 and 2) and 1 (toxicity grade 3 and 4) in both basic and extended datasets and adjusted odds ratios and 95% confidence interval of age, gender and 14 genetic variables was assessed.

Results

Univariate analysis

In the basic model, a total of 96 ALL subjects (70 boys and 26 girls) were included. The mean age of the patients was 9.8 ± 5.0 year. Among this cohort, hematological toxicity was observed as follows: grade 1 ($n = 35$), grade 2 ($n = 25$), grade 3 ($n = 29$) and grade 4 ($n = 7$). TPMT*3C, TYMT*12 (rs2842934), TPMT 12356 C>T (rs4449636), TPMT 16606 T>C (rs2518463), TPMT 26354 G>T (rs2842949), ITPA 94 C>A (rs1127354), ITPA 138 G>A (rs8362) and ITPA intron 2 (rs7270101) were identified by sequencing in this cohort. No other rare variants were observed in the studied samples. As ITPA intron 2 variant was observed only in one subject and hence excluded from the analysis. Logistic regression analysis revealed the independent associations of TPMT*12 (adjusted OR 3.69, 95% CI 1.12–12.17, $p = 0.03$) polymorphism with 6-MP-mediated hematological toxicity in ALL. The association of TPMT*3C with toxicity (adjusted OR 10.58, 95% CI 0.88–127.16, $p = 0.06$) was found to be marginal due to a relatively less frequency of this variant in our cohort (Table 1).

Toxicity prediction model

As shown in the CART model (Fig. 1), TPMT*12 (root node) was the main predictor of toxicity followed by TPMT*3C. The other determinants of toxicity were: TPMT int3, TPMT int4, TPMT int 7 and ITPA exon 3 variants, which inflate the toxicity when co-segregated in combination or with key determinants. This model had an accuracy of 93.6% with the area under the curve being 0.9649 (Fig. 2).

In both genders, TPMT (*12 and *3C) variant alleles were strongly associated with 6-MP-induced hematological toxicity with boys and girls exhibiting 8.67-fold (95% CI 2.56–29.32, $p < 0.001$) and 12.00-fold (1.73–83.46) risk, respectively.

Discussion

In the current study, we have employed the machine learning tools for the translation of demographic and pharmacogenetic data into hematological toxicity grade, a useful

Table 1 Logistic regression analysis showing the independent association of each variable with hematological toxicity

Variable	MAF	OR	95% CI	*p* value
Age	ND	0.97	0.88–1.07	0.53
Gender	ND	2.01	0.62–6.53	0.25
TPMT*3C (rs1142345)	3.1%	10.58	0.88–127.16	0.06
TPMT*12 (rs2842934)	28.6%	3.69	1.12–12.17	0.03*
TPMT int3 (rs4449636)	13.3%	0.45	0.06–3.29	0.43
TPMT int4 (rs2518463)	5.1%	12.6	0.55–288.64	0.11
TPMT int7 (rs2842949)	5.1%	Inf	0.00 – Inf	1.00
ITPA ex2 (rs1127354)	1.0%	0.00	0.00-Inf	1.00
ITPA ex3 (rs8362)	31.6%	0.71	0.24–2.04	0.52

MAF minor allele frequency, *OR* odds ratio, *CI* confidence interval, *ND* not determined

*Statistically significant

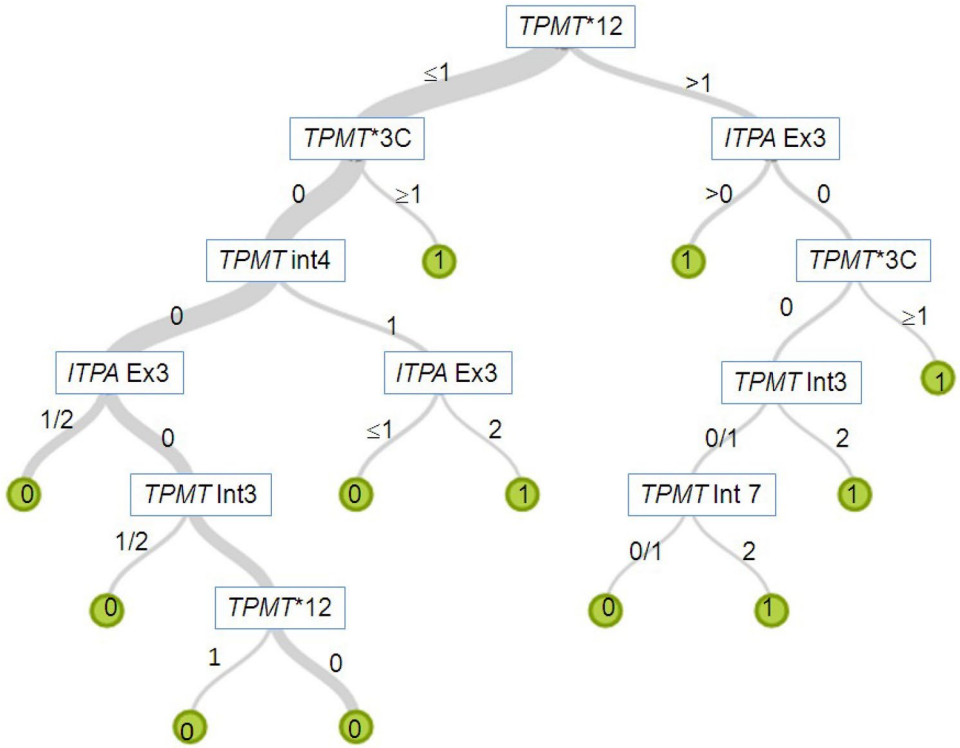

Fig. 1 Basic model of toxicity prediction. This classification and regression tree (CART) model was based on the data of 96 acute lymphoblastic leukemia patients on TPMT and ITPA variants. The genotypes namely wild, heterozygous and homozygous mutants were computed as 0, 1 and 2, respectively. The '0' and '1' with green background represent "low (grade 1 and 2)" and "high (grade 3 and 4)" toxicity

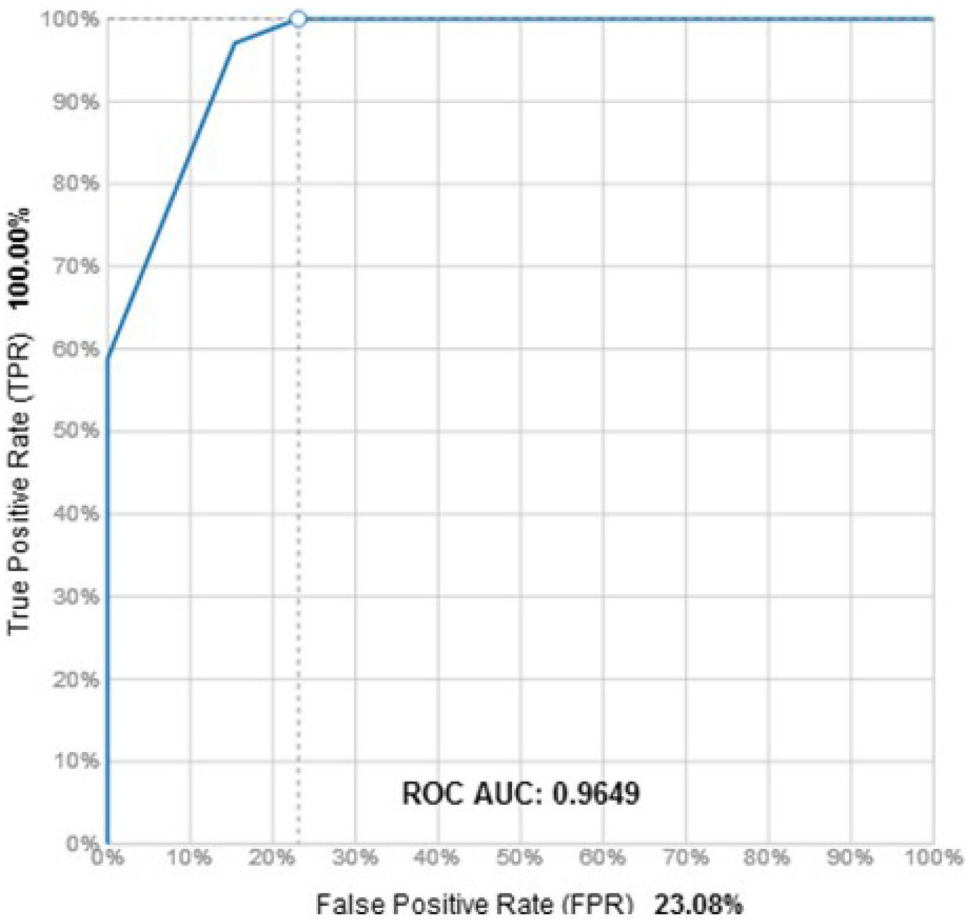

Fig. 2 Receiver operating characteristic (ROC) curve to assess clinical utility of prediction. The area under the ROC curve for predicted vs. observed toxicity was 0.9649, suggesting excellent predictability

clinical entity in tailoring the 6-MP dose in ALL patients. Such algorithm was used earlier for the differentiation of clinical responders and non-responders to thiopurine therapy in inflammatory bowel disease [8], which outperformed the metabolite-based predictions. The model developed in the current study highlights TPMT*3C and TPMT*12 as the key determinants of toxicity irrespective of gender.

The variant frequency of TPMT*12 was higher than that of TPMT*3C (28.6% vs. 3.1%), which emphasizes TPMT*12 variant allele as the most significant pharmacogenetic determinant of hematological toxicity. The frequency of TPMT*3C in our population was comparable to the Chinese population (3.1% vs. 3.6%) and our results are in concordance with their results in depicting the association of this variant with hematological toxicity [9]. TPMT*12 variant (Ser125Leu) was reported to have a significant decrease in intrinsic clearance of 6-MP with 30% residual TPMT activity [10], thus contributing to toxicity. ITPA rs8362 was reported to be associated with higher 6-TGN levels [11].

Apart from TPMT and ITPA, nucleoside diphosphate-linked moiety X-type motif 15 (NUDT15) rs116855232 [12], ATP Binding Cassette Subfamily C Member 4 (ABCC4) rs3765534 [13] and Protein Kinase C And Casein Kinase Substrate In Neurons 2 (PACSIN2) rs2413739 [14] variants were also reported to contribute towards 6-MP-mediated toxicity. Future studies incorporating these variants in predictions can improve the precision in predicting the toxicity. Advances in genomics coupled with machine learning have huge translational potential in ensuring safety and efficacy of therapeutic drug monitoring as depicted in the current study. The Clinical Pharmacogenetics Implementation Consortium (CPIC) formed guidelines on dose adjustment of 6-MP based on the genotype, which ensures the clinical translational of this information for the safety of the patients.

To conclude, TPMT*3C and TPMT*12 are the key determinants of 6-MP-mediated hematological toxicity in ALL patients. The developed prediction model has the potential in tailoring 6-MP therapy. The prediction model showed better performance characteristics in terms of accuracy, sensitivity and specificity in predicting the toxicity.

Acknowledgements We acknowledge the team of Department of Medical Oncology, Nizam's Institute of Medical Sciences, Hyderabad for their cooperation for this study.

Author contributions The study was conceived and designed by DR and VKK. The research and analytical work was carried out by SMN, PD, SAA and TH. The statistical models were developed and manuscript was drafted by SMN and VKK.

Funding No funding was received for this work.

Data availability The datasets during and/or analysed during the current study are available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Gerbek T, Ebbesen M, Nersting J, Frandsen TL, Appell ML, Schmiegelow K (2018) Role of TPMT and ITPA variants in mercaptopurine disposition. Cancer Chemother Pharmacol 81(3):579–586
2. El-Rashedy FH, Ragab SM, Dawood AA, Temraz SA (2015) Clinical implication of thiopurine methyltransferase polymorphism in children with acute lymphoblastic leukemia: a preliminary Egyptian study. Indian J Med Paediatr Oncol 36(4):265–270
3. Chrzanowska M, Kuehn M, Januszkiewicz-Lewandowska D, Kurzawski M, Droździk M (2012) Thiopurine *S*-methyltransferase phenotype-genotype correlation in children with acute lymphoblastic leukemia. Acta Pol Pharm 69(3):405–410
4. Adam de Beaumais T, Fakhoury M, Medard Y, Azougagh S, Zhang D, Yakouben K, Jacqz-Aigrain E (2011) Determinants of mercaptopurine toxicity in paediatric acute lymphoblastic leukemia maintenance therapy. Br J Clin Pharmacol 71(4):575–584
5. Dorababu P, Nagesh N, Linga VG, Gundeti S, Kutala VK, Reddanna P, Digumarti R (2012) Epistatic interactions between thiopurine methyltransferase (TPMT) and inosine triphosphate pyrophosphatase (ITPA) variations determine 6-mercaptopurine toxicity in Indian children with acute lymphoblastic leukemia. Eur J Clin Pharmacol 68(4):379–387
6. Dorababu P, Naushad SM, Linga VG, Gundeti S, Nagesh N, Kutala VK, Reddanna P, Digumarti R (2012) Genetic variants of thiopurine and folate metabolic pathways determine 6-MP-mediated hematological toxicity in childhood ALL. Pharmacogenomics 13(9):1001–1008
7. Ogungbenro K, Aarons L, CRESim, Epi-CRESim Project Groups (2015) Physiologically based pharmacokinetic model for 6-mercaptopurine: exploring the role of genetic polymorphism in TPMT enzyme activity. Br J Clin Pharmacol 80(1):86–100
8. Waljee AK, Joyce JC, Wang S, Saxena A, Hart M, Zhu J, Higgins PD (2010) Algorithms outperform metabolite tests in predicting response of patients with inflammatory bowel disease to thiopurines. Clin Gastroenterol Hepatol 8(2):143–150
9. Ma XL, Zhu P, Wu MY, Li ZG, Hu YM (2003) Relationship between single nucleotide polymorphisms in thiopurine methyltransferase gene and tolerance to thiopurines in acute leukemia. Zhonghua Er Ke Za Zhi 41(12):929–933
10. Hamdan-Khalil R, Allorge D, Lo-Guidice JM, Cauffiez C, Chevalier D, Spire C, Houdret N, Libersa C, Lhermitte M, Colombel JF, Gala JL, Broly F (2003) In vitro characterization of four novel non-functional variants of the thiopurine S-methyltransferase. Biochem Biophys Res Commun 309(4):1005–1010
11. Lee MN, Kang B, Choi SY, Kim MJ, Woo SY, Kim JW, Choe YH, Lee SY (2015) 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
12. Yi ES, Choi YB, Choi R, Lee NH, Lee JW, Yoo KH, Sung KW, Lee SY, Koo HH (2018) NUDT15 variants cause hematopoietic toxicity with low 6-TGN levels in children with acute lymphoblastic leukemia. Cancer Res Treat 50(3):872–882
13. Tanaka Y (2017) Susceptibility to 6-mercaptopurine toxicity related with NUDT15 and ABCC4 variants in Japanese childhood acute lymphoblastic leukemia. Rinsho Ketsueki 58(8):950–956
14. Smid A, Karas-Kuzelicki N, Jazbec J, Mlinaric-Rascan I (2016) PACSIN2 polymorphism is associated with thiopurine-induced hematological toxicity in children with acute lymphoblastic leukaemia undergoing maintenance therapy. Sci Rep 6:30244