HLA-DRB1*07:01–HLA-DQA1*02:01–HLA-DQB1*02:02 haplotype is associated with a high risk of asparaginase hypersensitivity in acute lymphoblastic leukemia

Nóra Kutszegi,[1,2] Xiaoqing Yang,[3] András Gézsi,[1] Géza Schermann,[1] Dániel J. Erdélyi,[2] Ágnes F. Semsei,[1] Krisztina M. Gábor,[4] Judit C. Sági,[1] Gábor T. Kovács,[2] András Falus,[1] Hongyun Zhang[5] and Csaba Szalai[1,6]

[1]Department of Genetics, Cell- and Immunobiology, Semmelweis University, Budapest, Hungary; [2]2nd Department of Paediatrics, Semmelweis University, Budapest, Hungary; [3]BGI-Shenzhen, China; [4]Department of Pediatrics and Pediatric Health Care Center, Faculty of Medicine, University of Szeged, Szeged, Hungary; [5]BGI Clinical Laboratory, Shenzhen, China and [6]Central Laboratory, Heim Pal Children Hospital, Budapest, Hungary

NK and XY contributed equally to this work.

Check the online version for the most updated information on this article, online supplements, and information on authorship & disclosures: www.haematologica.org/content/102/9/1578

ABSTRACT

Hypersensitivity reactions are the most frequent dose-limiting adverse reactions to *Escherichia coli*-derived asparaginase in pediatric acute lymphoblastic leukemia (ALL) patients. The aim of the present study was to identify associations between sequence-based Human Leukocyte Antigen Class II region alleles and asparaginase hypersensitivity in a Hungarian ALL population. Four-digit typing of *HLA-DRB1* and *HLA-DQB1* loci was performed in 359 pediatric ALL patients by using next-generation sequencing method. Based on genotypic data of the two loci, haplotype reconstruction was carried out. In order to investigate the possible role of the HLA-DQ complex, the *HLA-DQA1* alleles were also inferred. Multivariate logistic regression analysis and a Bayesian network-based approach were applied to identify relevant genetic risk factors of asparaginase hypersensitivity. Patients with *HLA-DRB1*07:01* and *HLA-DQB1*02:02* alleles had significantly higher risk of developing asparaginase hypersensitivity compared to non-carriers [$P=4.56 \times 10^{-5}$; OR=2.86 (1.73-4.75) and $P=1.85 \times 10^{-4}$; OR=2.99 (1.68-5.31); n=359, respectively]. After haplotype reconstruction, the *HLA-DRB1*07:01-HLA-DQB1*02:02* haplotype was associated with an increased risk. After inferring the *HLA-DQA1* alleles the *HLA-DRB1*07:01–HLA-DQA1*02:01–HLA-DQB1*02:02* haplotype was associated with the highest risk of asparaginase hypersensitivity [$P=1.22 \times 10^{-5}$; OR=5.00 (2.43-10.29); n=257]. Significantly fewer T-cell ALL patients carried the *HLA-DQB1*02:02* allele and the associated haplotype than did pre-B-cell ALL patients (6.5%; *vs.* 19.2%, respectively; $P=0.047$). In conclusion, we identified a haplotype in the Human Leukocyte Antigen Class II region associated with a higher risk of asparaginase hypersensitivity. Our results confirm that variations in HLA-D region might influence the development of asparaginase hypersensitivity.

Introduction

Asparaginase is a pivotal component of pediatric acute lymphoblastic leukemia (ALL) treatment. It converts asparagine to aspartic acid and ammonia. Lymphoblast cells lose the ability to produce asparagines; hence the loss of exogenous asparagine leads to cell death. However, the use of asparaginase can be challenging as either hypersensitivity reactions or neutralizing antibody formation can occur against this heterologous enzyme. Both can consequently lead to lower exposure to asparaginase which may result in suboptimal treatment response.[1-3] In

addition, hypersensitivity reactions to asparaginase range in severity from mild, transient (flushing or rash) to generalized anaphylaxis, which can be potentially life-threatening.[4]

Few studies have investigated the genetic basis of asparaginase hypersensitivity (AH). In a genome-wide association study (GWAS) of 485 pediatric ALL patients of European ancestry, Chen *et al.* found that rs4958351 of the glutamate receptor gene *GRIA1* located at 5q33 associated with AH.[5] Recently, in a GWAS of a multiethnic cohort of 3308 patients, Fernandez *et al.* replicated the finding between the rs4958351 variant and AH and found that the association was stronger in patients treated with native *Escherichia coli*-derived asparaginase (*E. coli* asparaginase) than in those receiving pegylated asparaginase.[6] However, the significance of the association (P=0.03) did not reach the genome-wide threshold (P=5 x 10[-8]). This association has been replicated in a relatively small Slovenian population of 146 patients.[7] In our previous candidate-gene study, we found that although genetic variants of *GRIA1* influenced the risk to AH, significant differences were observed according to sex and patient subgroup (T-cell or pre-B-cell ALL).[8]

The association of *Human Leukocyte Antigen (HLA)-DRB1*07:01* allele with an increased risk of AH was revealed in a candidate-gene study of Fernandez *et al.* of 1870 pediatric ALL patients of European ancestry.[9] Later, in the subsequent GWAS of the same group on AH, a SNP linked to *HLA-DRB1*07:01* also acted as a risk allele in patients of diverse ancestry. In this study, the rs6021191 variant in *NFATC2* was also associated with a higher risk of AH at the genome-wide significance threshold (P=4.1 x 10[-8], OR=3.11).[6] The minor allele frequency (MAF) of rs6021191 was only 0.001 among patients of European descent; therefore, this result has more relevance in patients of non-European ancestry.

HLA class II alleles are involved in presentation of peptides derived from extracellular proteins to T cells and subsequent activation of the immune response. They are located in the Major Histocompatibility Complex region on chromosome 6 and are expressed in antigen presenting cells. This genetic region is highly polymorphic with considerable linkage disequilibrium.

The primary goal of our study was to test the associations of HLA class II alleles with *E. coli* AH in a Hungarian population of 359 pediatric ALL patients using next-generation sequencing (NGS)-based HLA typing of *HLA-DRB1* and *HLA-DQB1* alleles. In addition, we aimed to evaluate the possible role of the *HLA-DRB1–HLA-DQA1–HLA-DQB1* haplotypes and the polymorphic amino acid positions located in the peptide-binding groove of the HLA-DQ complex in the mechanism of AH.

Methods

Patients

DNA samples from 359 pediatric ALL patients were available to investigate the association of HLA class II alleles with AH. The patients were treated with protocols from the Berlin-Frankfurt-Münster Study Group (accrued patients from 1990 to 2011): ALL-BFM 90 (n=72), ALL-BFM 95 (n=165), ALL IC-BFM 2002 (n=117), and ALL IC-BFM 2009 (n=5). The combined chemotherapy regimens contained *E. coli* asparaginase (Kidrolase or Asparaginase medac) as first-line treatment. The dosing schedules of *E. coli*

Table 1. Patients' characteristics.

Sex (%)	
Male	200 (55.7)
Female	159 (44.3)
Age at diagnosis, years	
Mean (± SD)	6.3 (±4.2)
Median (range)	4.8 (1-18)
Risk category (%)	
Low-risk	102 (28.4)
Medium-risk	213 (59.3)
High-risk	44 (12.3)
Immunophenotype (%)	
Pre-B ALL	287 (79.9)
T-ALL	46 (12.8)
Unknown	26 (7.2)

SD: Standard Deviation: pre-B ALL: pre-B-cell acute lymphoblastic leukemia (ALL); T-ALL: T-cell ALL.

asparaginase have been described in detail previously.[8]

Written informed consent was obtained from the study participants or from the next of kin, carers, or guardians on behalf of the minors/children who took part in the study. The study was conducted according to the Declaration of Helsinki and approved by the Hungarian Scientific and Research Ethics Committee of the Medical Research Council (ETT TUKEB; case n.: 8-374/2009-1018EKU 914/PI/08).

The patients had a median age of 4.8 years at diagnosis (range 1-18 years) (Table 1). The overall incidence of *E. coli* AH was 39.0%. Data collection was carried out retrospectively from medical records. The National Cancer Institute Common Toxicity Criteria (CTC) system v.3.0 was used to assess the grade of hypersensitivity. We regarded a case as AH when signs of allergic reactions or anaphylactic reactions CTC grade 1 and above were noted, as described earlier.[8]

Sequence-based typing of *HLA-DRB1* and *DQB1*

For sequencing, DNA samples from pediatric ALL patients with peripheral blood or bone marrow origin were available.[8] Only those patients whose DNA sample met the quality control criteria of NGS-based typing method were enrolled in this study. All of the 359 patients were typed at high-resolution level for *HLA-DRB1* and *HLA-DQB1*, in the Beijing Genomics Institute (BGI) using an NGS-based genotyping method that applies a massively parallel paired-end sequencing on Illumina MiSeq, effective variant phasing and haploid sequence assembling pipeline, as previously described.[10] Based on this, exon 2 was sequenced in *HLA-DRB1*, and exons 2 and 3 were sequenced in *HLA-DQB1* genes.

Estimation of *HLA-DRB1–HLA-DQA1–HLA-DQB1* haplotypes

HLA-DRB1–HLA-DQB1 haplotypes were estimated based on *HLA-DRB1* and *HLA-DQB1* 4-digit typing data of ALL patients by using the PHASE software (v.2.1.1). In addition, the *HLA-DRB1* and *HLA-DQB1* genotypic data was used to search for three-gene *HLA-DRB1–HLA-DQA1–HLA-DQB1* haplotypes in the Allele Frequency Net Database-HLA Haplotype Frequency Search (*http://www.allelefrequencies.net/hla6003a.asp*).[13] Using this approach, it was possible to infer independently the most likely haplotypes (including information about the *HLA-DQA1* locus) using haplotype data pertaining to Caucasoid ethnic origin as reference panel (*Online Supplementary Text S1*).

Inference of polymorphic amino acid positions in HLA class II alleles

Amino acid sequences encoded by the *HLA-DRB1, HLA-DQA1* and *HLA-DQB1* exon 2 were inferred using the four-digit sequencing results of *HLA-DRB1* and *HLA-DQB1* as well as the inference results for *HLA-DQA1*. The Alignment Viewer of Database of Major Histocompatibility Complex (*dbMHC;* https://www.ncbi.nlm.nih.gov/gv/mhc/align.fcgi?cmd=aligndisplay&user _id=0&probe_id=0&source_id=0&locus_id=0&locus_group=1&proto _id=0&kit_id=0&banner=1) was used to assess the polymorphic amino acid positions in HLA-DRB1, HLA-DQB1 and HLA-DQA1 chains.

Statistical analysis

*Frequentist methods*: multivariate logistic regression was used to test the associations of HLA class II alleles, *HLA-DRB1–HLA-DQA1–HLA-DQB1* haplotypes and polymorphic amino acid positions with *E. coli* AH. Sex, ALL immunophenotype (pre-B or T-ALL), age at diagnosis (≤10 or >10 years), risk group (standard- or medium- or high-risk), and treatment protocol were included in the model as categorical covariates. Assuming an additive genetic model, odds ratios (ORs) and 95% confidence intervals (CIs) were obtained to estimate risks for each variable to AH. To account for multiple testing Bonferroni correction was used ($P \leq 2.66 \times 10^{-4}$, based on a total of 66 HLA class II alleles, 38 haplotypes and 84 polymorphic amino acid positions tested). R statistical software (3.1.2) and IBM SPSS Statistic (v.20) software were used for analyses.

BN-BMLA method: we also applied Bayesian network based Bayesian multilevel analysis of relevance (BN-BMLA) to extend our genetic association study by estimating *a posteriori* probabilities of strong relevance (posteriors).[14-17] The BN-BMLA method was explained in detail in our previous studies.[18-20] Briefly, it computes the *a posteriori* probability of the strongly relevant variable sets with respect to a target variable (e.g. a dichotomous variable (yes/no) describing the AH status of the patients). The strongly relevant variables have a direct influence on the target. Values for posterior (P) range from 0 to 1, where P=1 and P=0 mean that the probability of the target being dependent on a predictor (e.g. HLA allele) is 100% and 0%, respectively. There were two types of strong relevance: direct relevance (e.g. casual genetic variant) and pure interaction (e.g. epistasis). For analyses, *HLA-DRB1, HLA-DQB1* and *HLA-DQA1* alleles, *HLA-DRB1–HLA-DQA1–HLA-DQB1* haplotypes, sex, ALL immunophenotype, age at diagnosis, risk group, and treatment protocol were included in the model as predictors.

Results

Association of *HLA-DRB1* and *HLA-DQB1* alleles with asparaginase hypersensitivity

We investigated the associations of HLA class II alleles (*HLA-DRB1, HLA-DQB1* and *HLA-DQA1*) and haplotypes with hypersensitivity reactions to *E. coli* asparaginase in
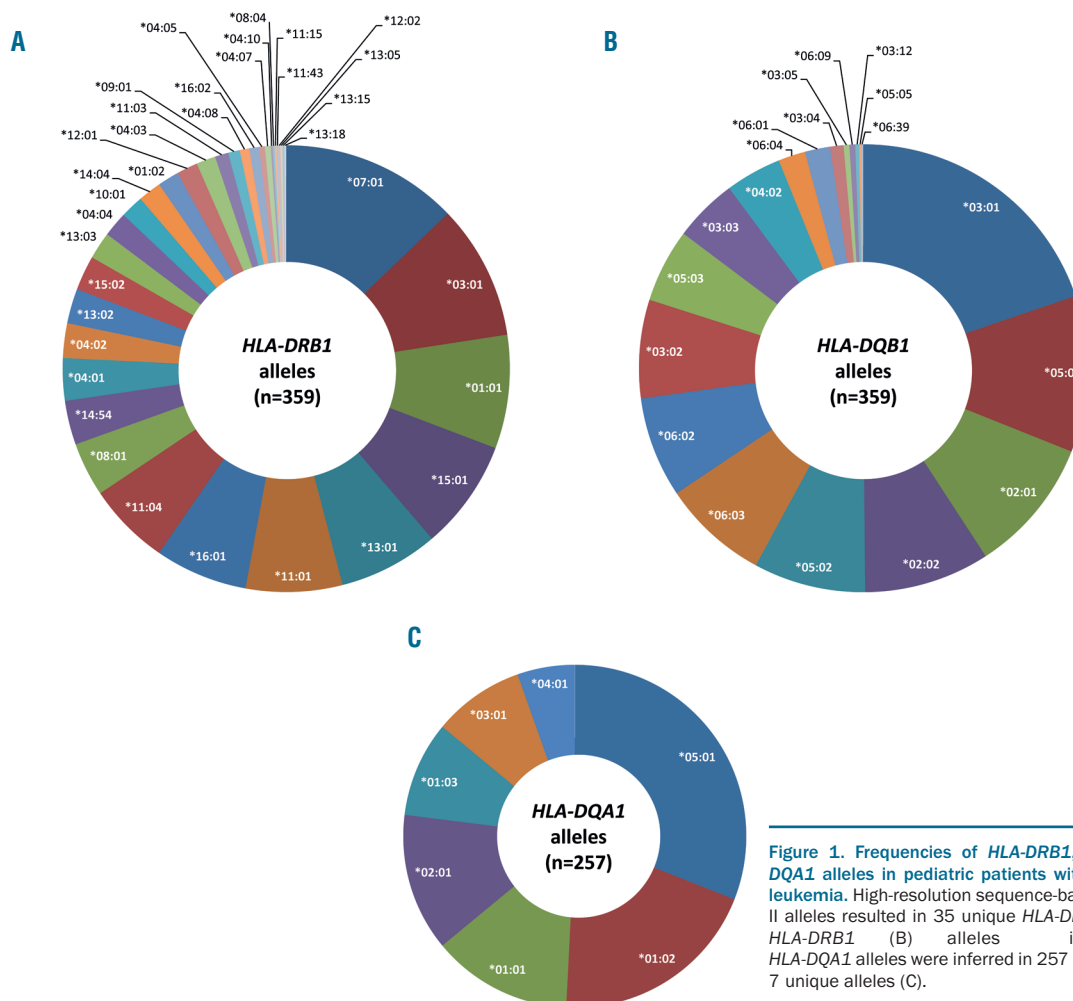

Figure 1. Frequencies of *HLA-DRB1, HLA-DQB1* and *HLA-DQA1* alleles in pediatric patients with acute lymphoblastic leukemia. High-resolution sequence-based typing of HLA class II alleles resulted in 35 unique *HLA-DRB1* (A) and 19 unique *HLA-DRB1* (B) alleles in 359 patients. *HLA-DQA1* alleles were inferred in 257 patients and resulted in 7 unique alleles (C).

Table 2. Association of *HLA-DRB1–HLA-DQB1* haplotypes with *E. coli* asparaginase hypersensitivity in acute lymphoblastic leukemia patients (n=359).

Haplotype	Frequency	OR[1]	95% CI[2]	P[1]
DRB1*01:01–DQB1*05:01	0.081	0.71	0.38-1.31	0.27
DRB1*01:01–DQB1*06:03	0.001	NA[3]	NA[3]	0.98
DRB1*01:02–DQB1*05:01	0.015	1.39	0.39-4.94	0.61
DRB1*03:01–DQB1*02:01	0.097	1.05	0.60-1.82	0.87
DRB1*04:01–DQB1*03:01	0.014	0.37	0.08-1.80	0.22
DRB1*04:01–DQB1*03:02	0.015	1.30	0.38-4.47	0.68
DRB1*04:01–DQB1*03:05	0.001	NA[3]	NA[3]	0.98
DRB1*04:02–DQB1*03:02	0.025	0.86	0.31-2.35	0.77
DRB1*04:03–DQB1*03:02	0.008	2.25	0.37-13.73	0.38
DRB1*04:03–DQB1*03:04	0.003	NA[3]	NA[3]	0.98
DRB1*04:03–DQB1*03:05	0.003	NA[3]	NA[3]	0.98
DRB1*04:04–DQB1*03:02	0.017	4.74	1.35-16.61	0.02
DRB1*04:04–DQB1*04:02	0.001	NA[3]	NA3	0.98
DRB1*04:05–DQB1*03:02	0.004	NA[3]	NA3	0.99
DRB1*04:07–DQB1*03:01	0.004	1.35	0.12-15.78	0.81
DRB1*04:08–DQB1*03:04	0.007	2.47	0.39-15.55	0.34
DRB1*04:10–DQB1*04:02	0.001	NA[3]	NA[3]	0.98
DRB1*07:01–DQB1*02:02	*0.091*	*2.99*	*1.68-5.31*	*1.85 x 10[-4]*
DRB1*07:01–DQB1*03:03	0.038	1.71	0.76-3.84	0.20
DRB1*08:01–DQB1*03:01	0.001	NA[3]	NA[3]	0.98
DRB1*08:01–DQB1*03:02	0.001	NA[3]	NA[3]	0.98
DRB1*08:01–DQB1*04:02	0.036	0.48	0.18-1.26	0.13
DRB1*08:04–DQB1*04:02	0.001	NA[3]	NA[3]	0.98
DRB1*09:01–DQB1*03:03	0.008	1.63	0.31-8.64	0.57
DRB1*10:01–DQB1*05:01	0.017	0.12	0.02-0.98	0.05
DRB1*11:01–DQB1*03:01	0.070	0.40	0.19-0.82	0.01
DRB1*11:03–DQB1*03:01	0.010	1.13	0.24-5.32	0.88
DRB1*11:04–DQB1*03:01	0.060	1.30	0.67-2.54	0.44
DRB1*11:15–DQB1*03:01	0.001	NA[3]	NA[3]	0.98
DRB1*11:43–DQB1*03:01	0.001	NA[3]	NA[3]	0.98
DRB1*12:01–DQB1*03:01	0.013	2.34	0.60-9.12	0.22
DRB1*12:01–DQB1*03:12	0.003	NA[3]	NA3	0.98
DRB1*12:02–DQB1*03:01	0.001	NA[3]	NA[3]	0.98
DRB1*13:01–DQB1*06:02	0.001	NA[3]	NA[3]	0.98
DRB1*13:01–DQB1*06:03	0.071	0.51	0.26-1.00	0.05
DRB1*13:02–DQB1*06:03	0.001	NA[3]	NA[3]	0.98
DRB1*13:02–DQB1*06:04	0.019	0.90	0.28-2.90	0.85
DRB1*13:02–DQB1*06:09	0.004	0.81	0.07-9.13	0.86
DRB1*13:03–DQB1*03:01	0.019	0.61	0.18-2.06	0.43
DRB1*13:05–DQB1*03:01	0.001	NA[3]	NA[3]	0.98
DRB1*13:15–DQB1*03:01	0.001	NA[3]	NA[3]	0.98
DRB1*13:18–DQB1*06:03	0.001	NA[3]	NA[3]	0.98
DRB1*14:04–DQB1*05:03	0.017	1.23	0.38-4.01	0.73
DRB1*14:54–DQB1*05:02	0.004	1.23	0.37-48.73	0.25
DRB1*14:54–DQB1*05:03	0.028	0.72	0.26-1.97	0.52
DRB1*15:01–DQB1*05:02	0.006	0.49	0.05-5.17	0.55
DRB1*15:01–DQB1*06:01	0.001	NA[3]	NA[3]	0.98
DRB1*15:01–DQB1*06:02	0.070	1.32	0.70-2.47	0.39
DRB1*15:01–DQB1*06:03	0.001	NA[3]	NA[3]	0.98
DRB1*15:01–DQB1*06:39	0.001	NA[3]	NA[3]	0.98
DRB1*15:02–DQB1*05:03	0.008	NA[3]	NA[3]	0.98
DRB1*15:02–DQB1*06:01	0.017	0.22	0.05-1.06	0.06
DRB1*16:01–DQB1*05:02	0.064	1.25	0.66-2.38	0.50
DRB1*16:01–DQB1*05:05	0.001	NA[3]	NA[3]	0.98
DRB1*16:01–DQB1*06:02	0.001	NA[3]	NA[3]	0.98
DRB1*16:02–DQB1*05:02	0.007	0.45	0.05-4.10	0.48

Results that reached the significance threshold (*P*≤2.66 x 10[-4]) are in bold. NA: not available. [1]For the association between the allele and asparaginase hypersensitivity. [2]95% confidence interval for the estimated odds ratio. [3]The odds ratios and confidence intervals were not estimated due to low allele frequency..
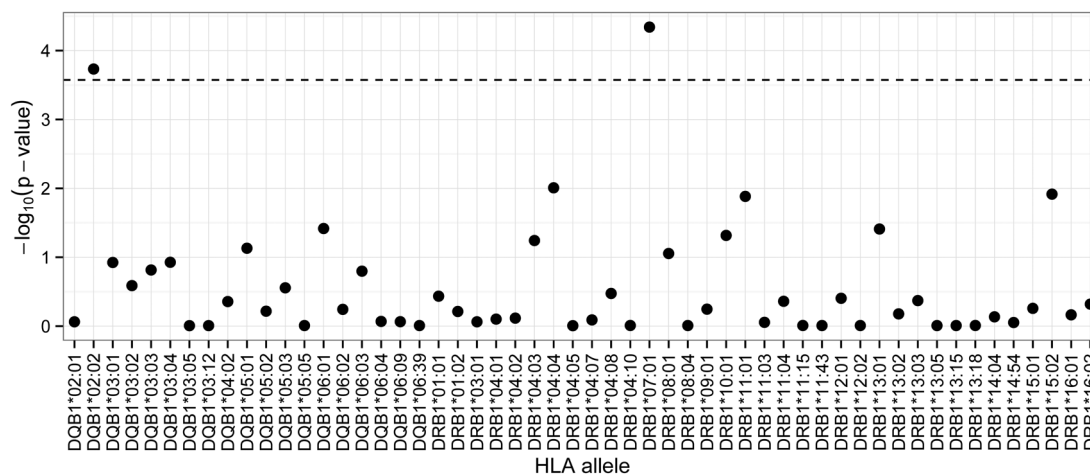
Figure 2. Association of *HLA-DQB1* and *HLA-DQB1* alleles with *E. coli* asparaginase hypersensitivity. Multivariate logistic regression was applied to assess the association of HLA class II alleles with asparaginase hypersensitivity. Sex, age, treatment protocol, risk group, acute lymphoblastic leukemia (ALL) immunophenotype were included in the analysis as categorical covariates. The dashed horizontal line indicates the Bonferroni-corrected statistical significance threshold. *HLA-DRB1*07:01* and *HLA-DQB1*02:02* alleles were significantly associated with asparaginase hypersensitivity (n=359; P=4.56x10^{-5} and 1.85x10^{-4}, respectively).

pediatric ALL patients. Among 359 patients, the high-resolution sequence-based typing resulted in 35 unique *HLA-DRB1* and 19 unique *HLA-DQB1* alleles (Figure 1A and B).

Univariate and multivariate logistic regression was performed to test the potential risk factors for AH (*Online Supplementary Table S1*). The incidence of AH varied by risk group (P=5.7 x 10^{-5}), ranging from 31% for medium-risk group to 77% for the high-risk group and treatment protocol (P=3.36 x 10^{-3}), ranging from 26% for ALL-BFM 90 to 60% for ALL IC-BFM 2002 experimental protocol (*Online Supplementary Table S1*). Using multivariate analysis risk group and treatment protocol also showed significant association with AH (*Online Supplementary Table S1*). For analyses with genetic data, we included all risk factors in the model as categorical covariates.

Applying Bonferroni correction (P≤2.66x10^{-4}) in multivariate logistic regression analyses *HLA-DRB1*07:01* and *HLA-DQB1*02:02* alleles showed significant associations with *E. coli* AH in 359 pediatric ALL patients (Figure 2). None of the patients was homozygous for either allele. Patients with *HLA-DRB1*07:01* allele had significantly higher risk of developing AH compared to non-carriers [P=4.56x10^{-5}; OR=2.86 (1.73-4.75)] (Figure 3). Similarly, patients harboring *HLA-DQB1*02:02* were at greater risk of having AH [P=1.85x10^{-4}; OR=2.99 (1.68-5.31)] (Figure 3). Allelic odds ratios and the AH rate of patients depending on the carrier state are shown in *Online Supplementary Table S2*.

There was a significant difference in the proportion of *HLA-DQB1*02:02* carriers between patients with pre-B-cell and T-cell ALL (19.2% *vs.* 6.5%; in patients with pre-B-cell and T-cell ALL, respectively; P=0.047). No such difference could be detected in the case of *HLA-DRB1*07:01* carriers.
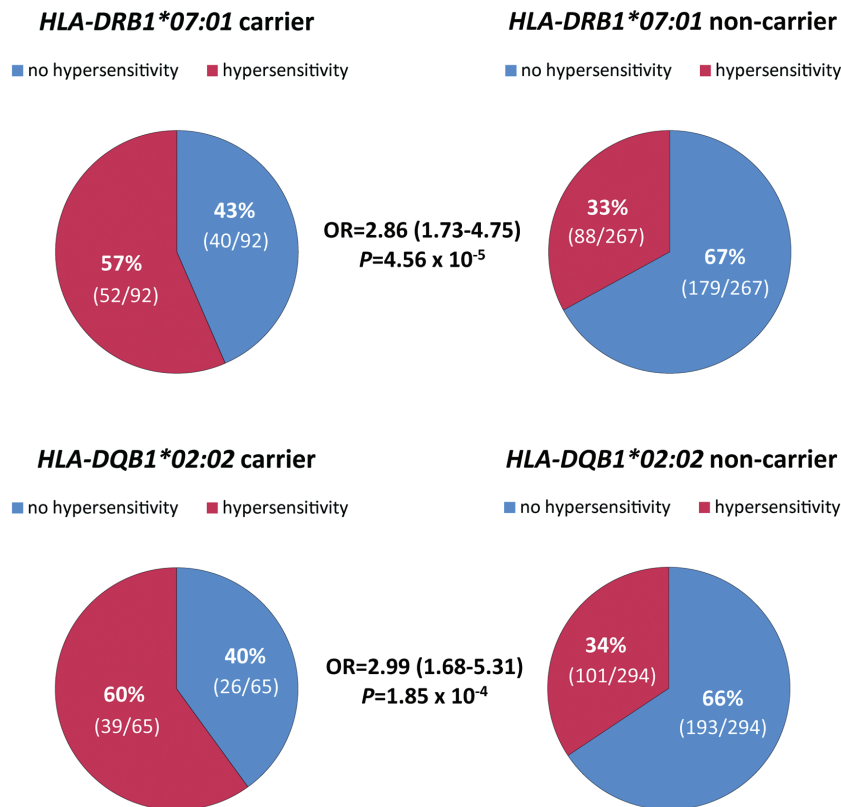
The BN-BMLA approach was used to investigate the different dependency relations between HLA alleles, age, ALL immunophenotype, sex, treatment protocol, risk group, and AH. The analysis revealed that if both *HLA-DRB1* and *HLA-DQB1* alleles were included in the Bayesian Network, among genetic factors only *HLA-DQB1*02:02* allele had direct strong relevance to AH (P=0.95) beside the influence of the risk group status (P=1.00) (Figure 4). There was no interaction or redundancy between predictors.

Haplotype estimation and the inference of *HLA-DQA1* alleles

In order to further investigate the relationship between *HLA-DRB1*07:01* and *HLA-DQB1*02:02* alleles associated with AH, we estimated haplotypes using the PHASE software (v.2.1.1). Using this method, the haplotype reconstruction resulted in 56 different haplotypes (Table 2). Two haplotypes containing *HLA-DRB1*07:01* allele were estimated among patients: *HLA-DRB1*07:01–HLA-DQB1*02:02* and *HLA-DRB1*07:01–HLA-DQB1*03:03*. Out of these, only *HLA-DRB1*07:01–HLA-DQB1*02:02* showed a positive association with AH [P=1.85 x 10^{-4}; OR=2.99 (1.68-5.31)] (Table 2). According to this, the *HLA-DRB1*07:01* and *HLA-DQB1*02:02* are risk alleles and in contrast to the *HLA-DRB1*07:01–HLA-DQB1*03:03* haplotype, *HLA-DRB1*07:01–HLA-DQB1*02:02* reached statistical significance. This is in agreement with the result of the BN-BMLA, i.e. in respect of AH, the presence of *HLA-DQB1*02:02* seems to be important.

Next, the *HLA-DRB1-HLA-DQA1-HLA-DQB1* haplotypes were independently inferred by using reference data from The Allele Frequency Net Database. The aim of inferring *HLA-DQA1* alleles was to investigate the possible role of the structure of the HLA-DQ complex in the pathomechanism of AH. In contrast to the HLA-DR complex, the alpha chain is also polymorphic in the HLA-DQ heterodimer, therefore polymorphic amino acid positions in HLA-DQA1 may also influence the peptide preference of the HLA-DQ complexes and consequently the develop-

HLA-DRB1*07:01 carrier

■ no hypersensitivity ■ hypersensitivity

HLA-DRB1*07:01 non-carrier

■ no hypersensitivity ■ hypersensitivity

OR=2.86 (1.73-4.75)
P=4.56 x 10^{-5}

HLA-DQB1*02:02 carrier

■ no hypersensitivity ■ hypersensitivity

HLA-DQB1*02:02 non-carrier

■ no hypersensitivity ■ hypersensitivity

OR=2.99 (1.68-5.31)
P=1.85 x 10^{-4}

Figure 3. Patients carrying the *HLA-DRB1*07:01* or *HLA-DQB1*02:02* alleles had strong association with asparaginase hypersensitivity. The incidence of hypersensitivity reactions was 57% (52 of 92) and 33% (88 of 267) for patients carrying *HLA-DRB1*07:01* allele and for non-carriers, respectively. In the case of patients with *HLA-DQB1*02:02* allele, the incidence of asparaginase hypersensitivity was 60% (39 of 65), while it was 34% (101 of 294) for patients who did not have the allele. Odds ratios (OR) and 95% confidence intervals were estimated by using multivariate logistic regression analysis.

ment of a specific immune response against asparaginase. Using data for haplotype frequencies in Caucasoid populations, and applying the aforementioned method, the estimation of extended haplotypes was possible in 72% of the patients (n=257; *Online Supplementary Table S3*). We compared the haplotype estimation results obtained by using the PHASE software to those extended haplotype data inferred by using The Allele Frequency Net Database. Results were in 100% agreement. Finally, among these 257 patients, 7 unique *HLA-DQA1* alleles were inferred (Figure 1C).

Multivariate logistic regression analysis showed that *HLA-DQA1*02:01* allele and *HLA-DRB1*07:01–HLA-DQA1*02:01–HLA-DQB1*02:02* haplotype were positively and significantly associated with AH [*P*=3.45x10^{-5}; OR=3.69 (1.99-6.84) and *P*=1.22x10^{-5}; OR=5.00 (2.43-10.29), respectively] (*Online Supplementary Table S3*). These statistical associations were equivalent to the associations of *HLA-DRB1*07:01* and *HLA-DQB1*02:02* alleles, respectively; hence *HLA-DQA1*02:01* exclusively and in all cases occurred together with *HLA-DRB1*07:01*.

Inference of polymorphic amino acid positions in HLA class II alleles

A total of 84 polymorphic amino acid positions were identified from 4-digit typing results of *HLA-DRB1* and *HLA-DQB1* genes and from inferred *HLA-DQA1* allelic data using dbMHC database. In order to further investigate the possible sequence features that confer a greater risk to AH, multivariate logistic regression was performed to test for associations between polymorphic amino acid positions in HLA class II alleles and AH. In our patients

with leukemia, a total of 27 positions associated significantly (Table 3).

A valine at position 78 in HLA-DRB1 showed the strongest association with AH [*P*=8.16x10^{-6}; OR=4.01 (2.18-7.37)] (Table 3). Out of the 27 risk amino acids, 7 and 3 were uniquely coded by *HLA-DRB1*07:01* and *HLA-DQA1*02:01*, respectively [*P*=3.45x10^{-5}; OR=3.69 (1.99-6.84)] (Table 3). All positions with risk amino acids in HLA-DQB1 belonged to at least two alleles and were, to a lesser extent, associated with hypersensitivity [*P*=1.69x10^{-4}; OR=2.87 (1.66-4.97)] (Table 3).

Interestingly, while all amino acids conferring risk to AH were present in *HLA-DQB1*02:02*; the other *HLA-DQB1* allele linked to *HLA-DRB1*07:01*, *HLA-DQB1*03:03* possessed no risk amino acid at any risk position. These results confirm the important role of the *HLA-DRB1*07:01–HLA-DQA1*02:01–HLA-DQB1*02:02* haplotype in the increased risk of AH.

Discussion

The aim of our study was to investigate the association of the variations in the HLA class II region with *E. coli* AH in pediatric ALL patients. By analyzing 359 Hungarian patients, we found that *HLA-DRB1*07:01* and *HLA-DQB1*02:02* were associated with an increased risk of developing AH. Haplotype reconstruction also revealed that the *HLA-DRB1*07:01–HLA-DQA1*02:01–HLA-DQB1*02:02* haplotype positively and significantly associated with an increased risk of AH.

Previously, in 2014, a study by Fernandez *et al.* of 1870

Table 3. Polymorphic amino acid positions in HLA-DRB1, HLA-DQB1 and HLA-DQA1 associated with *E. coli* asparaginase hypersensitivity in acute lymphoblastic leukemia patients (n=257).

HLA class II locus	AA position[1]	Risk residue[2]	OR[3]	95% CI[4]	P[3]	Other possible amino acids[5]	HLA class II alleles with risk residue
HLA-DRB1	10	Q	2.22	1.45-3.40	2.30×10^4	E,Y	*01:01, *01:02, *04:01, *04:02, *04:03, *04:04, *04:05, *04:07, ***07:01**, *09:01, *15:01, *15:02,*16:01, *16:02
	11	G	3.69	1.99-6.84	3.45×10^5	D,L,P,S,V	*07:01
	13	Y	3.69	1.99-6.84	3.45×10^5	F,G,H,R,S	*07:01
	14	K	3.69	1.99-6.84	3.45×10^5	E	*07:01
	25	Q	3.69	1.99-6.84	3.45×10^5	R	*07:01
	30	L	3.69	1.99-6.84	3.45×10^5	C,G,H,R,Y	*07:01
	37	F	3.69	1.99-6.84	3.45×10^5	L,N,S,Y	*07:01
	57	V	3.52	2.00-6.19	1.31×10^5	D,S	*07:01, *09:01, *12:01
	60	S	3.52	2.00-6.19	1.31×10^5	Y	*07:01, *09:01, *12:01
	73	G	2.96	1.71-5.12	1.10×10^4	A	*07:01, *03:01
	74	Q	3.69	1.99-6.84	3.45×10^5	A,E,L,R	*07:01
	78	V	4.01	2.18-7.37	8.16×10^6	Y	*07:01, *09:01
HLA-DQA1	47	K	3.69	1.99-6.84	3.45×10^5	C,Q,R	*02:01
	50	L	3.03	1.78-5.14	4.09×10^5	E,V	*02:01, *03:01
	52	H	3.69	1.99-6.84	3.45×10^5	R,S	*02:01
	53	R	3.03	1.78-5.14	4.09×10^5	K,Q	*02:01, *03:01
	54	L	3.69	1.99-6.84	3.45×10^5	F	*02:01
HLA-DQB1	28	S	2.87	1.66-4.97	1.69×10^4	T	*02:01, *02:02
	30	S	2.87	1.66-4.97	1.69×10^4	H,Y	*02:01, *02:02
	37	I	2.87	1.66-4.97	1.69×10^4	D,Y	*02:01, *02:02
	46	E	2.87	1.66-4.97	1.69×10^4	V	*02:01, *02:02
	47	F	2.87	1.66-4.97	1.69×10^4	Y	*02:01, *02:02
	52	L	2.87	1.66-4.97	1.69×10^4	P	*02:01, *02:02
	55	L	2.87	1.66-4.97	1.69×10^4	P,R	*02:01, *02:02
	57	A	2.60	1.58-4.26	1.61×10^4	D,S,V	*02:01, *02:02, *03:02, *03:04, *03:05
	71	K	2.87	1.66-4.97	1.69×10^4	A,D,T	*02:01, *02:02
	74	A	2.87	1.66-4.97	1.69×10^4	E,S	*02:01, *02:02

HLA class II alleles significantly associated with asparaginase hypersensitivity are in bold. [1]Amino acid (AA) position of the HLA class II protein sequence without the leader signal sequence. [2]Amino acid associated with asparaginase hypersensitivity. [3]For the association between the allele and asparaginase hypersensitivity. [4]95% confidence interval for the estimated odds ratio. [5]Other possible amino acids refer to other amino acids within that amino acid position observed in our patients (n=257).

pediatric ALL patients of European ancestry revealed the association of *HLA-DRB1*07:01* allele with AH.[9] Later, the same group confirmed the role of this allele in an ethnically diverse pediatric population. In our study, we also confirmed these findings, but after haplotype reconstruction we also found that only the *HLA-DRB1*07:01–HLA-DQB1*02:02* haplotype associated significantly, and the *HLA-DRB1*07:01–HLA-DQB1*03:03* haplotype did not reach statistical significance, possibly because of the lower allele frequency (0.038 *vs.* 0.091 for the *HLA-DRB1*07:01–HLA-DQB1*02:02* haplotype). However, further investigations are warranted to elucidate the actual causality. The importance of the *HLA-DQB1*02:02* allele in the hypersensitivity was also shown by the BN-BMLA method, which showed that if both *HLA-DRB1* and *HLA-DQB1* alleles were included in the Bayesian Network only *HLA-DQB1*02:02* allele had direct strong relevance to AH with a high posterior probability (0.95). The situation is the same with the *HLA-DQA1*02:01* allele as with the *HLA-DRB1*07:01* allele, because they are in complete linkage disequilibrium with each other. This may suggest that, in

our population, the *HLA-DQB1*02:02* allele also seems to be an important genetic risk factor for AH in this region.

However, when the amino acid results are evaluated, there is some controversy in this statement. It is well known that proteins coded by the HLA-D, or also known as the MHC class II genes play a critical role in the immune response to extracellular antigens. Antigen presenting cells present processed antigens, or epitopes bound in a pocket formed by the MHC II chains to T-cell receptors on the surface of helper T cells, resulting in cytokine secretion by T cells leading to, among others, differentiation of B cells into antibody-secreting plasma cells. According to the theory, the stronger the MHC II protein binds an epitope, the stronger the immune response is. The strength of the binding is determined by the 3D structure of the MHC II protein which depends on its amino acid composition. Table 2 shows the amino acids associated with the AH and the different HLA class II alleles coding for these variants. Regarding *E. coli* asparaginase epitopes, Fernandez *et al.* predicted with bioinformatic methods that the *HLA-DRB1*07:01* is a high binding allele
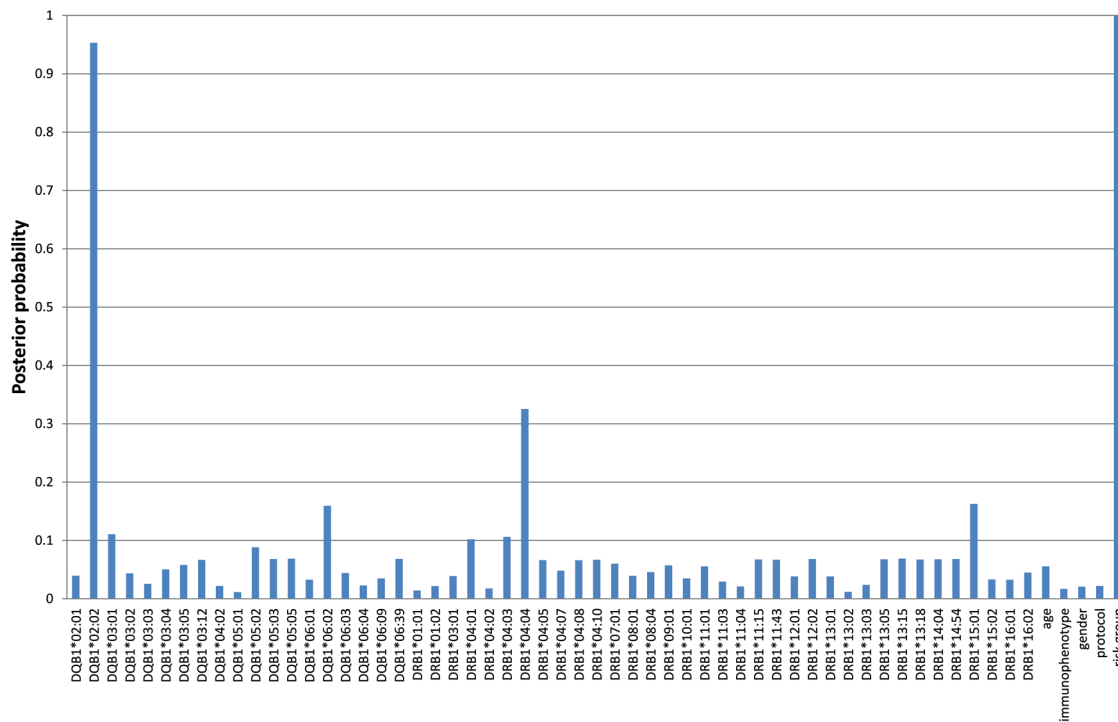
Figure 4. *A posteriori* **strong relevance of predictors to asparaginase hypersensitivity.** The BN-BMLA approach was used to investigate the different dependency relations between HLA alleles, age, acute lymphoblastic leukemia (ALL) immunophenotype, sex, treatment protocol, risk group and asparaginase hypersensitivity. *HLA-DQB1*02:02* allele and risk group had direct strong relevance to asparaginase hypersensitivity (*P*=0.95 and 1.00, respectively).

which could be the reason for its association with AH. In our results, several amino acids within the HLA-DQB1 protein were also associated with increased AH, and all of them were present in *HLA-DQB1*02:02*. But they were also present in *HLA-DQB1*02:01* which was not associated with AH. It means that from these amino acid results it cannot be predicted why amino acids in *HLA-DQB1*02:02* were associated with AH. This may mean that although the presence of *HLA-DQB1*02:02* allele may be decisive in the development of AH, the presence of the *HLA-DRB1*07:01* and/or *HLA-DQA1*02:01* alleles may also be necessary for the increased risk, raising the possibility that it is the extended haplotype rather than the individual alleles that is important in this respect.

There was another interesting finding in the present study. A significantly smaller proportion of T-cell ALL patients carried the *HLA-DQB1*02:02* allele than did pre-B-cell ALL patients (6.5%; *vs.* 19.2% of patients with T-cell and pre-B-cell ALL, respectively). This may suggest two interpretations. Either the allele increases the risk to pre-B-cell ALL or decreases the risk to T-cell ALL. Earlier the *HLA-DQB1*02:02* allele was found to be associated with increased risk of celiac disease indicating its autoimmunogenic potential.[21] Although we did not determine its allele frequency in a healthy control population, its frequency data in The Allele Frequency Net Database in Caucasian populations were around the value found in the pre-B-cell ALL patients in our study. Thus, in contrast its role in celiac disease, the *HLA-DQB1*02:02* allele might decrease the risk of the development of T-cell ALL.

Furthermore, because this allele is always on the *HLA-DRB1*07:01–HLA-DQA1*02:01–HLA-DQB1*02:02* haplotype, this could mean that this haplotype might provide some protection against development of T-cell ALL but not against pre-B-cell ALL. Naturally, subjects from the same population must be compared to test this assumption.

This study has some limitations. First, patients who died during the chemotherapy due to therapy-resistant progressive disease, or due to infections or toxicities of therapy are under-represented in our ALL population. Furthermore, data about hypersensitivity reactions to *E. coli* asparaginase was collected retrospectively from the patients' files. This does not allow for meticulous documentation or fine grading of hypersensitivity reactions.

In some cases the inference of *HLA-DQA1* alleles could not be precisely carried out. For further analyses with *HLA-DQA1* alleles, data from undetermined haplotypes were merged into one of the possible *HLA-DQA1* allele categories (*Online Supplementary Table S4*). In most cases, exon 2 of the involved alleles was identical at the amino acid sequence level and no haplotype was found with the other *HLA-DQA1* allele in The Allele Frequency Net Database either. In the case of *HLA-DRB1*13:01–HLA-DQA1*01:02/*01:03–HLA-DQB1*06:0*3 haplotype, the *HLA-DQA1* allele was regarded as *HLA-DQA1*01:0*3 because it was more likely based on the overall Caucasian haplotype frequency data. To increase the reliability of our inference results, a strict algorithm was applied, which led to the over-representation of common haplotypes.

There are, however, several strengths to this study. First, we studied a relatively large and homogenous European population (Hungarian). Second, we used NGS technology to determine the *HLA-DQB1* and *HLA-DQB1* HLA types, which is more accurate than the SNP-based imputation used in other studies. In addition, the nucleotide sequence-based HLA typing, in contrast to the costly and laborious direct experimental HLA typing, allows polymorphic HLA amino acid positions from the HLA alleles to be inferred. Third, ours is the first study to evaluate the role of HLA haplotypes in AH, and our results indicated that the HLA haplotype-based risk prediction might be more precise than the single allele-based risk calculation.

In conclusion, in our study, we found that the *HLA-DRB1*07:01–HLA-DQA1*02:01–HLA-DQB1*02:02* haplotype was associated with a high risk of *E. coli* AH in pediatric ALL patients. Based on our results, besides the previously described *HLA-DRB1*07:01* allele, the *HLA-DQB1*02:02* allele might also be important in the development of AH. This needs further investigation.

References

1. Silverman LB, Gelber RD, Dalton VK, et al. Improved outcome for children with acute lymphoblastic leukemia: results of Dana-Farber Consortium Protocol 91-01. Blood. 2001;97(5):1211-1218.
2. Rizzari C, Conter V, Stary J, et al. Optimizing asparaginase therapy for acute lymphoblastic leukemia. Curr Opin Oncol. 2013;25 Suppl 1:S1-9.
3. Pieters R, Hunger SP, Boos J, et al. L-asparaginase treatment in acute lymphoblastic leukemia: a focus on Erwinia asparaginase. Cancer. 2011;117(2):238-249.
4. Schmiegelow K, Attarbaschi A, Barzilai S, et al. Consensus definitions of 14 severe acute toxic effects for childhood lymphoblastic leukaemia treatment: a Delphi consensus. Lancet Oncol. 2016;17(6):e231-239.
5. Chen SH, Pei D, Yang W, et al. Genetic variations in GRIA1 on chromosome 5q33 related to asparaginase hypersensitivity. Clin Pharmacol Ther. 2010;88(2):191-196.
6. Fernandez CA, Smith C, Yang W, et al. Genome-wide analysis links NFATC2 with asparaginase hypersensitivity. Blood. 2015;126(1):69-75.
7. Rajic V, Debeljak M, Goricar K, Jazbec J. Polymorphisms in GRIA1 gene are a risk factor for asparaginase hypersensitivity during the treatment of childhood acute lymphoblastic leukemia. Leuk Lymphoma. 2015:1-6.
8. Kutszegi N, Semsei AF, Gezsi A, et al. Subgroups of Paediatric Acute Lymphoblastic Leukaemia Might Differ Significantly in Genetic Predisposition to Asparaginase Hypersensitivity. PLoS One. 2015;10(10):e0140136.
9. Fernandez CA, Smith C, Yang W, et al. HLA-DRB1*07:01 is associated with a higher risk of asparaginase allergies. Blood. 2014;124(8):1266-1276.
10. Cao H, Wang Y, Zhang W, et al. A short-read multiplex sequencing method for reliable, cost-effective and high-throughput genotyping in large-scale studies. Hum Mutat. 2013;34(12):1715-1720.
11. Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet. 2005;76(3):449-462.
12. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 2001;68(4):978-989.
13. Gonzalez-Galarza FF, Takeshita LY, Santos EJ, et al. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. Nucleic Acids Res. 2015;43(Database issue):D784-788.
14. Hullám G, Antal P, Szalai C, Falus A. Evaluation of a Bayesian model-based approach in GA studies. In: Sašo D, Pierre G, Juho R, eds. Proceedings of the third International Workshop on Machine Learning in Systems Biology. Proceedings of Machine Learning Research: PMLR, 2009:30-43.
15. Antal P, Hullam G, Gézsi A, Millinghoffer A. Learning complex bayesian network features for classification. In: Studený M, Vomlel J, editors. Proceedings of the Third European Workshop on Probabilistic Graphical Models; 2006 September 12–15; Prague: Czech Republic: Action M Agency; 2006. p. 9-16.
16. Antal P, Millinghoffer A, Hullám G, et al. Bayesian, systems-based, multilevel analysis of associations for complex phenotypes: from interpretation to decisions. In: Sinoquet C, Mourad R, eds. Probabilistic Graphical Models for Genetics, Genomics and Postgenomics. Oxford: Oxford University Press, 2014:318-352.
17. Antal P, Millinghoffer A, Hullám G, Szalai C, Falus A. A Bayesian View of Challenges in Feature Selection: Feature Aggregation, Multiple Targets, Redundancy and Interaction. In: Yvan S, Huan L, Iñaki I, Louis W, Yves Van de P, eds. Proceedings of the Workshop on New Challenges for Feature Selection in Data Mining and Knowledge Discovery at ECML/PKDD 2008. Proceedings of Machine Learning Research: PMLR, 2008:74-89.
18. Ungvari I, Hullam G, Antal P, et al. Evaluation of a partial genome screening of two asthma susceptibility regions using bayesian network based bayesian multilevel analysis of relevance. PLoS One. 2012;7(3):e33573.
19. Lautner-Csorba O, Gezsi A, Erdelyi DJ, et al. Roles of genetic polymorphisms in the folate pathway in childhood acute lymphoblastic leukemia evaluated by Bayesian relevance and effect size analysis. PLoS One. 2013;8(8):e69843.
20. Lautner-Csorba O, Gezsi A, Semsei AF, et al. Candidate gene association study in pediatric acute lymphoblastic leukemia evaluated by Bayesian network based Bayesian multilevel analysis of relevance. BMC Med Genomics. 2012;5:42.
21. Dieli-Crimi R, Cenit MC, Nunez C. The genetics of celiac disease: A comprehensive review of clinical implications. J Autoimmun. 2015;64:26-41.