Rīgas Stradiņa universitāte

Madara Kreile

# GENETIC FACTORS IN THE ETIOLOGY OF CHILDHOOD ACUTE B CELL PROGENITOR LYMPHOBLASTIC LEUKEMIA

Summary of Doctoral Thesis for obtaining the degree of a Doctor of Medicine

Subfield − Paediatrics

Riga, 2015

Doctoral Thesis was carried out at the Rīga Stradiņš University, Children's Clinical University Hospital

Scientific supervisor:

*Dr. med.* **Baiba Lace**,
Latvian Biomedical Research and Study centre

Reviewers:

*Dr. habil. med.*, Professor **Valda Staņēviča**,
Rīga Stradiņš University, Latvia

*Dr. biol.*, **Zane Kalniņa**,
Latvian Biomedical Research and Study centre

*Dr. med.*, **Andrejs Ivanovs**,
University of Edinburgh, United Kingdom

Defence of the Doctoral Thesis will take place at the public session of the Doctoral Committee of Medicine on 15 October at 15.15 in Hippocrates Lecture Theatre, 16 Dzirciema Street, Rīga Stradiņš University.

The Doctoral Thesis is available at RSU Library and RSU homepage: www.rsu.lv

Secretary of the Doctoral Committee:

*Dr. med.* **Jana Pavāre**

# TABLE OF CONTENTS

# ABBREVIATIONS

ALL – acute lymphoblastic leukaemia
ARID5B – AT-rich interactive domain 5B coding gene
bp – base pair
CI – confidence interval
DN – dominant negative
DNS – desoxyrhibandcleic acid
ESE – exon splicing *enhancers*
GWAS – genom wide association studies
*IKZF1* – Ikaros family zinc finger protein 1
IL – interleukin
LD – *linkage disequilibrium*
MAF – minor allele frequency
Mb – megabase
*MDR1* – P glucoprotein coding gene
MĢZL – Scientific laboratory of medical genetic
MLPA – multiplex ligation-dependent probe amplification
MRD – minimal residual disease
MTHFR – methylen tetrahydrofolate reductase
OR – odds ratio
PCR – ollymerase chain reaction
$r^2$ – correlation coefficient
RR – relative risk
SNP – single-nucleotide alelic variation
$\chi^2$ – chi square

# 1. INTRODUCTION

Hematopoesis is a multi-stage process, which results in development of blood and immande system's cells. Ikaros, which is coded by *IKZF1* and Ikaros faimily zinc finger transcription factors have a cricitcal role in haematopoesis, mostly in securing the differentiation, homeostasis and fandction of lymphoid cells (*Payne et al.*, 2011). Mice with heterozygotic mutation in Ikaros protein coding gene at the age from three to six months, in a result of loss of heterozigoty developed leukaemia and lymphoma, which were not combatible with life (*Winandy et al.*, 1995).

Acute lymphoblastic leukaemia (ALL) is one of the most frequent malignities in childhood (*Handger et al.*, 2013). The most frequent ALL subtype is pre-B cell ALL, and it forms approximately 80% of all ALL cases (*Urayama et al.*, 2013). The peak of case usually is seen between two and five years of age (*Seghatoleslam et al.*, 2012). Annual ALL incidince is 1 – 4.75 cases per 100 000 up to 15 years of age (*Redaelli et al.*, 2005). ALL more frequently is seen in boys, the calculated frequency ratio between boys and girls in the USA is 1.35 : 1 (*Siegel et al.*, 2014).

Despite wide research, the etiology of ALL is still andclear (*Nousome et al.*, 2013). Less than 5% of ALL is connected to genetic syndromes (*Pui et al.*, 2008).

It is believed, that development of acute lymphoblastic leukaemia is connected with genetic damage in T or B lymphocyte precursor cells, which then leads to their andlimited self-dupliation and/or genetic changes, which then lead to stopping of differentiation (*Mullighan*, 2012).

Scientists have identified more that 50 regions, in which changes in numbers of DNA copies can occur. Approximately in 60% of pre-B cell ALL cases the changes are in genes *PAX5, IKZF1, EBF1* and *LEF1*, which are invlovled in normal development of B lymphocytes (*Mullighan et al.*, 2007). In

*PAX5* gene mutations are foandd in 32% of ALL cases (*Woo et al.*, 2014). Mutation Gly183Ser in a heterozygotic state predisposes to development of leukaemia (*Shah et al.*, 2013).

There have been several genome wide association studies for identifying the possible genetic risk factors. In connection with higher ALL riski n wide association study two allelic variations were identified rs10821936 and rs10994982, which are localized in *ARID5B* gene, (*Trevino et al.*, 2009). Alelic variation rs4132601, which is localized in gene's *IKZF1* 3' non-translateable region and alelic variation rs2239633, which is localized in *CEBPE* has been shown to have close relationship to the risk of development of ALL (*Papaemmanuil et al.*, 2009).

Connecting the wide range studies of genome, one more alelic variation, which is connected with higher risk of ALL development was identified – rs3731217, which is localized in the gene's *CDKN2A* 1st introne (Sherborne et al., 2010).

There are still ongoing research about the possible ALL connection to xenobiothic metabolism. One of the metabolic pathways, which could play a role in the development of acute lymphoblastic leukaemia, is the matbolic pathway of folate, because it takes part in the synthesis, repairment and methylation of DNA (*Lupo et al.*, 2012). The protein, coded by *MTHFR* gene has a significant role in the folate metabolism, its allelic variations rs1801133 and rs1801131 decreases the activity of the enzyme (*Jain et al.*, 2012; *Yousefian et al.*, 2014), which can cause hypometilation of DNS, thus resulting in higher expression and activization in pro-oncogenes. (*J Yan et al.*, 2012).

Gene *MDR1* codes P glycoproteine, one of its fandctions is to protect the organism against xenobiotics, which might have a mutagenous activity (*Semsei et al.*, 2008). Allelic variations rs1045642 and rs2032582 in a homozigotic state decreases the activity of the enzyme (*Llaudo et al.*, 2013; *Sterjev et al.*, 2012) When the enzyme activity is dedreased, also the ability of glycoprotein

P to excrete toxic xenobiotcis and environment cancerogenes from the cells is decreased (*Wang et al.*, 2012).

Individuals with deletion in the genes that code glutathion S transferase *GSTT1* and *GSTM1* does not have one or both enzyme activity (*Dandna et al.*, 2013), which determinēs, that they cannot effectively excrete carcinogenes (*LY Xu et al.*, 2014), more often they develop somatic mutations and DNA fragments, which have a covalent bind with chemical substances (*Q Tang et al.*, 2013).

Gene *NQO1* codes NAD(P)H dehydrogenase quinone 1, if singe nucleotide polymorphism (SNP) rs1800566 is in a homozygotic state, the enzyme is almost inactive (*Misra et al.*, 2000). NQO1 fandction is to reducē and detoxify quinones and their derivates to protect cells from oxydative stress and cancerogenesis. SNP, which have a direct impact to enzyme activity, can predispose to tumour development (*Yang et al.*, 2015).

The producē tof the *IL15* gene is interleukine 15 (IL15), which is a pleyothrope cytokine (*Williams et al.*, 2014), which ha san impact on normal T and B lymphocyte and neutrophil proliferation, growth and differentiation (*Agostini et al.*, 1997; *Fehniger et al.*, 2001). A study, in which five SNP, loalized in gene *IL15* revealded, that SNP rs10519612 and rs17007695 have a satistically significant connection to increased risk of development of leukaemia in adults (*D Lin et al.*, 2010).

More than 20 possible environmental risk factors have been described, which could have an impact to the risk of developing leukaemia, but only a few of these results have been replicated or have a biological significance. Epidemiologic studies have proven, that there is a correlation between infections and the risk of ALL development. (*Inaba et al.*, 2013), and also between jonizing radiation and the risk of ALL development (*Belson et al.*, 2007). Among the risk factors of acute lymphoblastic leukaemia are also high birth weight (*Milne et al.*, 2013). ALL risk has been associatiod with maternal

facotrs, e.g., history oof stillbirth, which could be ralted to predispostion or environmental impact, also with increased materanl age, in this case the possible pathogenetic mechanism is the lack of chromosomal splitting in meiosis, also DNA damage caused by other facctors (*Mejia-Arangure et al.*, 2003).

## 1.1. Hypothesis

Genetic variations in genes, connected with differentiation of leukocytes, inluding the *IKZF1* genes, and in those, which are connected with metabolism of xenobiotics, have impact of development and course of acute lymphoblastic leukaemia in childhood.

## 1.2. Aim of the study

To find out geneti markers and their significance in development and course of acute lymphoblasti leukaemia, by using analysis of molecular genetics data, immandohistochemical examination and analysis of patients' risk groups.

## 1.3. Tasks of the study

1. Form a study group and collet a peripheral blood sample from individuals, which have been diagnosed with acute pre-B-cell leukaemia andder 18 years of age in from janury 2005 to july 2014, and also collect peripheral blood samples from their parents.

2. Perform an analysis of genetic risks on included individuals to evaluate the impact of allelic variations to the course of ALL.

3. Perform the replication of wide genome research data in Latvian population by analysing the abovementioned single nucleiotide allelic variation detection in genes *IKZF1, ARID5B,* CDKN2A and *CEBPE*, also detection of

previously not described allelic variations, which are localized in the *ARID5B* gene's third introne, using case control association model, family association model and hybrid-methode, which includes both above metioned methods.

4. Determine the connection of possible risk of ALL development with described signel nucleotide allelic variations: genes involved in xenobiotoc metabolism – *MDR1*, *MTHFR*, *NQO1*, *GSTT1* and *GSTM1*, also allelic variations localized in *IL15* gene and mutation in the *PAX5* gene.

5. Perform the full sequencing of *IKZF1* gene in all individuals and in the only patient, who at the dime of diagnosis was yoandger than one year of age also MLPA – deletion analysis. Determine the protein expression coded by the gene in the bone marrow in the exacerbation of the leukaemia, and analyse the expression in peripheral blood in samples with low expression level and in samples from patients without bone marrow samples.

## 1.4. The scientific novity

The thesis summarizes data about the patients who are diagnosed with acute pre-B cell lymphoblastic leukemia in the period from January 2005 to Jande 2014 . This is the first study in Latvia in which in the patients with acute pre- B cell leukemia the genetic markers studied . This is the first research in which the full IKZF1 gene sequencing is performed in ALL patients, allelic variation role in non- somatic cells and the potential relevance of development acute pre-B cell leukemia risk is analysed, as well as protein expression analysis in bone marrow cells and peripheral blood .

## 1.5. Practical aspects

Thesis is more of a fanddamental research work, which is based on studying the etiology of acute pre-B cell leukaemia in Latvian population, which is connected with impact of genetic factors. By fully studying genetic

factors and other factors in ALL etiopatogenetics, it is possible, it can be fully explained, thus improving the anddersating of pathology.

In this research we identified allelic variations, which have correlation with higher acute lymphoblasti leukaemia development risk in children, and also those, which are connected to poorer prognosis or have a protective role.

# 2. MATERIALS AND METHODS

## 2.1. Materials

Patients, who were diagnosed with acute pre-B cell lymphoblastic leukamiea in the time period from 2005 to July 2014 in the haemathological andit of Andiversity Children's Hospital were included in tis study. Only those individuals who have been diagnosed andder 18 years of age were included. Prior to enrollment individuals signed an informed consent. If an individual was a minor at the time of biological material collection, one or both of the biological parents signed informed consent forms in accordance with Latvian Central Ethics Committee approval. Patient inclusion criteria can be seen in figure 2.1.

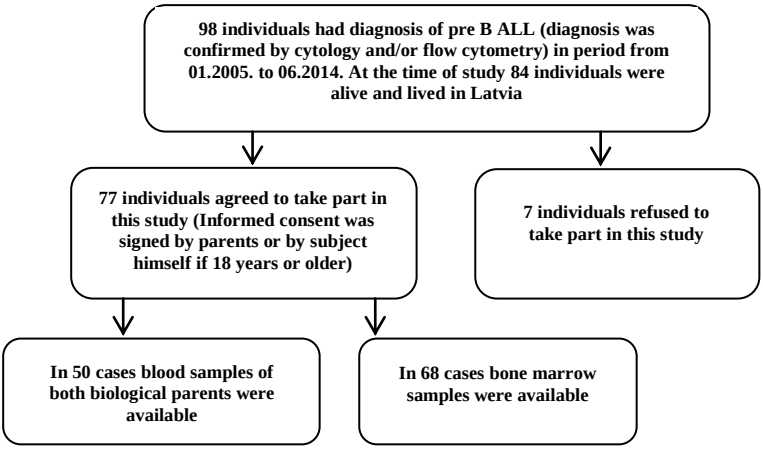


Fig. 2.1. **Patient inclusion criteria**

Control group was formed of 122 DNA samples of individuals of matched age and sex, that have been stored in the Scientific Laboratory of Medical Genetics (MĢZL) collection, and whose parents have signet the informed consent form at the time of the sample collection in which is stated,

that these samples can be used without limitations, accordingly to the researchers discretion. None of the control group individuals, which in July 2014 were yoandger than 18 years had diagnosis of acute pre-B cells lymphoblastic leukaemia.

## 2.2. Methods

### 2.2.1. Selection of genetic markers

Selection of genetic markers was performed based on publications, in this study for replication of other population' s data the following allelic variations of genes were selected: in the *ARID5B* gene – rs10821936, rs10821938, rs7089424, rs10994982, rs7923074, rs7896246; *IKZF1* gene – rs4132601; *CEBPE* – rs2239633; *CDKN2A* – rs3731217. For additional analysis two gene allelic variations were chosen, which were not analysed before, they were localised in the gene *ARID5B* 3$^{rd}$ introne – rs10821937 and rs7908445.

Allelic variations of the gene, that are localized in genes, included in xenobiotic metabolism genes, were selected by the most frequent allelic variations described in literature, which have the highest probability of connection with higher risk of leukaemia development. For the study the most sudiet allelic variations in genes *MDR1* (rs1045642 and rs2032582), *MTHFR* (rs1801131 and rs1801133) and *NQO1* (rs1800566) were analysed, also deletions in genes *GSTT1* and *GSTM1* were detected, despite the fact, that data from other studies are still not uniform.

For the first time in patients in childhood with ALL for analysis allelic variations in gene *IL15* (rs10519612, rs10519613, rs17007695) were chosen, previous publications describe their connection with higher risk of ALL in the adulthood.

Gene *PAX5* mutation NM_001280547.1:c.547G>A analysis was selected, because there have been publications, that this mutation in isolated heterozygotic state can predispose to development of leukaemia.

By the data published in literature, the full analysis of *IKZF1* sequence in children with ALL has not been done. Taking into account the *IKZF1* coded Ikaros protein function and significance of in differentiation of lymphocytes, the *IKZF1* was selected as the possible candidate gene in development of acute lymphoblastic leukaemia.

### 2.2.2. DNA extraction from venous blood sample

Venous blood was collected in ethylendiamintetraacethate vaccuteiner. DNA was extracted with standart fenole chloroform method, which is described by *John et al*., method is adapted in SLMG (*John et al.*, 1991).

### 2.2.3. Polymerase chain and restriction reaction fragment length allelic variation analysis

Performing amplification reactions the standart amplification reagent mixture was prepared. With restriction endonuclease analysed allelic variation syntethic oligonucleotide sequences were search in the program Praimer 3 (http://primer3.ut.ee/) or selected according to the described data in publications (*Alpman et al.*, 2010; *Ayaz et al.*, 2013; *Hanson et al.*, 2001; *Kim et al.*, 2006; *Kimura et al.*, 2005; *D Lin et al.*, 2010; *Safarinejad et al.*, 2012).

For detection of GSTT1 and GSTM1 null gentotypes in homozygotic state multiplex polymerase chain reaction was used, when in the same time deletion in both genes is detected. The method is adapted in SLMG from *Kondo* et al methodics (*Kondo et al.*, 2009).

### 2.2.4. Sequencing reaction

Samples for sequencing were prepared by the the manufacturer' s protocol. Electroferogramms were analysed, using the programme "*Chromas"* 2.4. The Acquired sequences were compared to *"BLAST" (Altschul et al., 1997)* available reference sequence. Sequencing raction was performed in seven *ARID5B* gene allelic variations. Full sequencing *IKZF1* gene was performed in all probands. The sixth and the eighth exone was sequenced in the control group patients (77 individuals, matched by age and sex).

### 2.2.5. Immunochemical analysis of bone marrow and blood samples

After bone marrow biopsy the sample was fixed in 10% formaline solution. The histologic cut was performed with a rotation microtome. Paraffine blocs and histological cuts were done in the Children's University Hospitals Histological laboratory of the Pathology unit.

The blood sample smear was made by transferring 1.5 microlitre blood from the patient in full remission to the microscope glass.

Immunochemical reactions were performed by the adapted protocol from the manufacturer protokola (*Dako*, USA).

### 2.2.6. Analysis of the blood sample for possible deletions in *IKZF1* gene

In the proband, which developed an acute pre-B cell lymphoblastic leukaemia under one year of age, we performed multiplex ligation-dependant probe amplification in the laboratory „*BioAnalytica Genotypus*" in Greece, using commercially available probe P202-B1 mixture (MRC, the Netherlands).

### 2.2.7. Statistical analysis

Study data was analysed, using descriptive and analytical statistical methods.

Statistical analysis in the control group and family study model for analysing the possible allelic variation connection with ALL wa performed, using PLINK 1.07 (*Purcell et al.*, 2007). GSTT1 and GSTM1 chi-square thest was done in JavaStat (*Rosner*, 2006).

For case control and family study model statistical result gathering, and also for increasing the statistical confidence in these results, a hybridanalysis was used, in which analysis was was performed, using programme R, *Haplin* add-on, which is based to a log-linear model, and if necessaru, uses – *expectation–maximization* EM algorithm for haploptype reconstruction (*Jugessur et al.*, 2009). Additionally, the allelic variation of the *MDR* gene rs2032582 analysis with adapted log-linear model was done in this program, because the allelic variation in tri-allelic.

For data statistical analysis of patients' morbidity age, protein expression in the bone marrow and its connection with gentypes and course of the illness, the programme *SPSS 20* (IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.) was used.

To analyse the identified allelic variation in *IKZF1* and its possible impact on splicing, the processing possible regions were analysed in *Human splicing finder* – HSF 3.0 (*Desmet et al.*, 2009).

For analysis of allelic variation disequilibre linkage (*linkage disequilibrium* – LD), the programmature *haploview* was used.

# 3. RESULTS

## 3.1. ALL patient descriptives

36 girls and 41 boys with ALL were included in this study. The boy to girl ratio in this study was 1.14 : 1.

The age at the time of the diagnostics of ALL was 0 to 17 years, the peak incidence was seen in the age group from 2 to 5 years of age. The mean age at the time of falling ill was 6.12 years, SD 4.7. The age distribution of patiens can be seen in the figure 3.1.

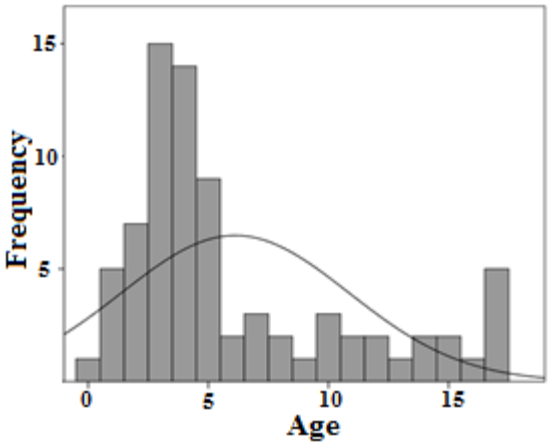


Fig. 3.1. **Age distribution in patients with ALL at the time of diagnostics**

The highest morbidity of ALL was seen among children born in 2004 – 14.3%. The number of patiens by the birth uear and the number of live-born children in the specific year is depicted in the figure 3.2. Data about the live-born children was acquired from the Central Statistical Bureau (www.csb.gov.lv).

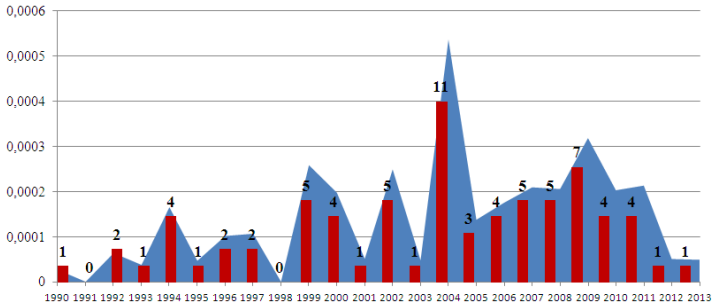


Fig. 3.2. **Division of ALL patients by the birth year**

Dividing patiens in the risk group, in which the BMF 05 and COG high risk criteria were combined, 48 patiens were in the standart risk group but 27 in the high risk group. Two patients were not included in the risk-group assessment, because one of them at the time of diagnosis was younger than one year of age, but other had the therapy initiated in Mexico, thus the leukocyte count at the time of diagnosis was not known.

Besides the prognosis worsening factors, also the prognostically positive factors were analysed, e.g., hyperploidy. Cytogenetic analysis was available in 64 individuals, 40% of them had hyperploidy in the karyotype.

During the time of this study 9 patients or 11.68% developed recidive, 5 patients (6.49%) died.

## 3.2. *ARID5B* gene's 3rd introne allelic variations' analysis

In total we analuzed 8 allelic variations, localized in the *ARID5B* gene's 3rd intron, whose MAF > 5%.

Besides that, also analysis for 2 allelic variations, that are in the sequenced region, rs77918077 and rs12246030, were analysed, they did not match the inclusion criteria. The allelic variation rs77918077 A allele was identified in 6 probands. The requency of A allele was 4.17%. In the control group the A allele frequency was 2.46 %. Between both gropus statisti-

cally significant differences were not found (p = 0.374 OR = 1.73 95% CI 0.48 – 6.17).

Based on the 1000 genome project first phase data, the allelic variation of rs12246030 rarely seen allele was G, which is seen in 3% of cases (*Abecasis et al.*, 2012). In this study G allele was not identified in any of individuals – neither study nor control group.

Performing allele association analysis isolatedly for each SNP, all allelic variation genotype distribution matched Hardy-Weinberg equilibrium.

Six of the eight analysed allelic variations in case control study were identified as statistically significant risk allele, which is connected ti higher risk of ALL development. Correcting this result by sex, it was detected, that the rarely seen allele was detected statistically significantly more frequently in boys than girls. The case control allele association analysis is depicted in the table 3.1.

Table 3.1.

***ARID5B* gene 3rd introne's allelic variation risk allele analysis in case control study**

| Gene's allelic variation | MAF | Frequency in patient group | Frequency in control group | p value | OR CI 95% | p value, corrected by sex |
|---|---|---|---|---|---|---|
| rs10994982 | A | 0.51 | 0.41 | 0.054 | 0.67 (0.44-1) | 0.059 |
| rs7908445 | T | 0.43 | 0.32 | 0.022 | 0.61 (0.40–0.93) | 0.03 |
| rs7923074 | A | 0.44 | 0.32 | 0.016 | 0.6 (0.39–0.91) | 0.022 |
| rs10821936 | C | 0.34 | 0.24 | 0.027 | 0.61 (0.39–0.95) | 0.042 |
| rs10821937 | C | 0.34 | 0.23 | 0.018 | 0.58 (0.37–0.91) | 0.03 |
| rs7896246 | A | 0.34 | 0.23 | 0.018 | 0.58 (0.37–0.91) | 0.03 |
| rs10821938 | A | 0.47 | 0.38 | 0.074 | 0.69 (0.46–1.04) | 0.07 |
| rs7089424 | G | 0.34 | 0.23 | 0.018 | 0.58 (0.37–0.91) | 0.03 |

MAF – minor allele frequency; OR – odds ratio; CI – confidence interval

When analysing the genotype connection with higher ALL development risk or each allelic variation seperately. for seven allelic variations statistically significant results were identified in the recesive hereditary model (analyzing gentype DD in comparison to Dd+dd). In the allelic model, comparing dominant ja recesive allele's frequencies, statistically significant results were seen in six allelic variations. These results can be seen in the table 3.2.

Table 3.2.

***ARID5B* gene 3rd introne's allelic variation risk allele analysis by hereditary model in in case control association study**

| Gene's allelic variation | MAF | $\chi^2$ for the recessive model | p value for the recessive model | $\chi^2$ for the allelic model | p value for the allelic model |
|---|---|---|---|---|---|
| rs10994982 | A | 2.2 | 0.13 | 3.72 | 0.05 |
| rs7908445 | T | 4.8 | 0.028 | 5.23 | 0.02 |
| rs7923074 | A | 4.8 | 0.028 | 5.84 | 0.016 |
| rs10821936 | C | 5.39 | 0.02 | 4.89 | 0.027 |
| rs10821937 | C | 5.33 | 0.02 | 5.58 | 0.018 |
| rs7896246 | A | 4.34 | 0.037 | 5.58 | 0.018 |
| rs10821938 | A | 5.94 | 0.015 | 3.19 | 0.07 |
| rs70894224 | G | 5.33 | 0.02 | 5.59 | 0.018 |

Evaluating the allelic variations' connection with risk of ALL development in the family study, all 8 analyzed allelic variations were statistically significantly related to ALL. Results are shown in table 3.3.

Table 3.3.

***ARID5B* gene 3rd introne's allelic variation risk allele analysis in family connection study**

| Gene's allelic variations | OR CI 95% | p value |
|---|---|---|
| rs7908445 | 2.53 (1.39−4.61) | 0.002 |
| rs7923074 | 2.79 (1.51−5.13) | $0.6 \times 10^{-3}$ |
| rs10821936 | 2.62 (1.38 −4.96) | 0.002 |
| rs10821937 | 3.18 (1.62−6.27) | $0.4 \times 10^{-3}$ |
| rs7896246 | 2.9 (1.41−5.95) | 0.002 |
| rs10821938 | 1.91 (1.12− 3.230 | 0.015 |
| rs7089424 | 3 (1.56−5.77) | $0.5 \times 10^{-3}$ |
| rs10994982 | 1.88 (1.05−3.39) | 0.032 |

For enforcing the statistical confidence of these results, both study models – case control and patient-parent trio study were unified by hybrid-method (*Jugessur et al.*, 2009). Results of this are shown in table 3.4.

Table 3.4.

***ARID5B* gene 3rd introne's allelic variation association hybridanalysis**

| SNP | One risk allele in genotype | | Two risk alleli in genotype | |
|---|---|---|---|---|
| | RR CI 95% | p value | RR CI 95% | p value |
| rs7908445 | 1.26 (0.72–2.21) | 0.416 | 3.35 (1.58–7.01) | 0.002 |
| rs7923074 | 1.37 (0.78–2.4) | 0.279 | 3.57 (1.67–7.55) | 0.001 |
| rs10821936 | 1.16 (0.66–2.04) | 0.6 | 4.61 (2.07–10.1) | $0.4 \times 10^{-3}$ |
| rs10821937 | 1.35 (0.76–2.35) | 0.312 | 5.29 (2.32–11.9) | $0.2 \times 10^{-3}$ |
| rs7896246 | 1.26 (0.72–2.18) | 0.43 | 3.96 (1.73–8.93) | 0.002 |
| rs10821938 | 1.21 (0.69–2.13) | 0.503 | 2.33 (1.08– 4.8) | 0.031 |
| rs7089424 | 1.32 (0.74–2.31) | 0.338 | 5.11 (2.23–11.4) | $0.2 \times 10^{-3}$ |
| rs10994982 | 1.3 (0.71–2.33) | 0.384 | 2.43 (1.14–5.13) | 0.024 |

SNP – single nucleotide polymorphism; RR – relative risk; CI – confidence interval

Evaluating the allelic variations linkage disequilibrium, located in the 3rd introne of *ARID5B* gene, and performing the LD analysis, the following result was acquired – see figure 3.3.

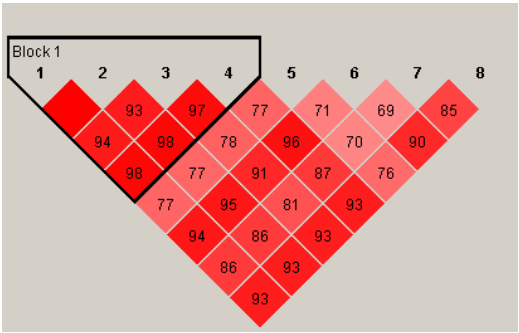


Fig. 3.3. **Analysis of linkage disequilibrium of the chromosome 10 gene *ARID5B* 3rd introne**

The allelic variations of the gene were localised as follows: 1 – rs7908445; 2 – rs7923074; 3 – rs10821936; 4 – rs10821937; 5 – rs7896246; 6 – rs10821938; 7 – rs10994982; 8 – rs7089424

The possible risk haplotypes wwere analysed in both study models – case control and family study. Haplotypes were formed from three allelic variations in the haploptype up to all 8 analyzed simoultaneously. The statistically significant results of haplotype analysis in the case-control study are shown in the table 3.5.

Table 3.5.

***ARID5B* gene allele variation haplotype analysis in case control association study**

| SNP included in the haplotype | Haplotype | Frequency in ALL patients | Frequency in control group | Chi square | p value |
|---|---|---|---|---|---|
| Three SNP | | | | | |
| 1/2/3 | ATA | 0.29 | 0.4 | 5.2 | 0.022 |
| 1/2/3 | GCC | 0.56 | 0.46 | 3.97 | 0.046 |
| 2/3/4 | TAC | 0.23 | 0.33 | 4.72 | 0.03 |
| 2/3/4 | CCT | 0.68 | 0.56 | 5.87 | 0.015 |
| 3/4/5 | ACC | 0.23 | 0.33 | 4.53 | 0.033 |
| 3/4/5 | CTT | 0.68 | 0.57 | 4.89 | 0.02 |
| 4/5/6 | CCA | 0.18 | 0.3 | 8.74 | 0.003 |
| 5/6/7 | CAA | 0.17 | 0.3 | 9.95 | 0.002 |
| 6/7/8 | AAG | 0.17 | 0.3 | 9.52 | 0.002 |
| Four SNP | | | | | |
| 1/2/3/ 4 | ATAC | 0.22 | 0.32 | 4.83 | 0.028 |
| 2/3/4/ 5 | TACC | 0.23 | 0.33 | 4.71 | 0.029 |
| 2/3/4/ 5 | CCTT | 0.68 | 0.58 | 4.54 | 0.033 |
| 3/4/5/ 6 | ACCA | 0.18 | 0.3 | 8.26 | 0.004 |
| 4/5/6/ 7 | CCAA | 0.17 | 0.3 | 8.77 | 0.003 |
| 5/6/7/ 8 | CAAG | 0.17 | 0.3 | 9.28 | 0.002 |
| FiveSNP | | | | | |
| 1/2/3/ 4/5 | ATACC | 0.24 | 0.31 | 4.68 | 0.031 |
| 2/3/4/ 5/ 6 | TACCA | 0.18 | 0.3 | 8.48 | 0.004 |
| 3/4/5/ 6/ 7 | ACCAA | 0.17 | 0.3 | 7.91 | 0.005 |
| 4/5/6/ 7/ 8 | CCAAG | 0.17 | 0.29 | 8.21 | 0.004 |
| Six SNP | | | | | |
| 1/2/3/ 4/5/ 6 | ATACCA | 0.17 | 0.29 | 7.29 | 0.007 |
| 2/3/4/ 5/ 6/7 | TACCAA | 0.17 | 0.3 | 8.13 | 0.004 |
| 3/4/5/ 6/7/ 8 | ACCAAG | 0.17 | 0.29 | 7.37 | 0.006 |
| Seven SNP | | | | | |
| 1/2/3/ 4/5/6/ 7 | ATACCAA | 0.17 | 0.28 | 6.63 | 0.009 |
| 2/3/ 4/ 5/6/7/ 8 | TACCAAG | 0.17 | 0.29 | 7.32 | 0.007 |
| Eight SNP | | | | | |
| 1/2/3/ 4/5/6/ 7/ 8 | ATACCAAG | 0.17 | 0.29 | 6.69 | 0.009 |

Allele variation description in the table: 1 – rs10994982, 2 – rs7908445, 3 – rs7923074, 4 – rs10821936, 5 – rs10821937, 6 – rs7896246, 7 – rs10821938 and 8 – rs7089424

But the family study model statistically significant haplotype was formed of sevena allele variations: rs7908445/rs7923074/ rs10821936/ rs10821937/rs7896246/rs10821938/ rs7089424 – TACCAAG (p value = 0.0008). The analysis of family haplotype is shown in the table 3.6.

Table 3.6.

***ARID5B* gene allele variation haplotype analysis in family association study**

| SNP included in the haplotype | Haplotype | Transmitted allele | Non-transmtted allel | Chi square | p value |
|---|---|---|---|---|---|
| 1/2/3 | ATA | 31 | 11 | 9.52 | 0.002 |
| 1/2/3 | GCC | 11 | 27 | 6.74 | 0.009 |
| 2/3/4 | TAC | 33.7 | 11 | 11.54 | $0.7 \times 10^{-3}$ |
| 2/3/4 | CCT | 13 | 37 | 11.52 | $0.7 \times 10^{-3}$ |
| 3/4/5 | ACC | 32.92 | 10 | 12.24 | $0.5 \times 10^{-3}$ |
| 3/4/5 | CTT | 12 | 35 | 11.26 | $0.8 \times 10^{-3}$ |
| 4/5/6 | CCA | 32.88 | 9.99 | 12.23 | $0.5 \times 10^{-3}$ |
| 4/5/6 | TTG | 12.12 | 30.08 | 7.65 | 0.006 |
| 5/6/7 | CAA | 35.88 | 10 | 14.6 | $0.1 \times 10^{-3}$ |
| 6/7/8 | AAG | 33.95 | 10 | 13.05 | $0.3 \times 10^{-3}$ |
| 1/2/3/ 4 | ATAC | 27.99 | 9.99 | 8.53 | 0.003 |
| 1/2/3/ 4 | GCCT | 10.99 | 26.99 | 6.74 | 0.009 |
| 2/3/4/ 5 | TACC | 32.92 | 10 | 12.24 | $0.5 \times 10^{-3}$ |
| 2/3/4/ 5 | CCTT | 12 | 34 | 10.52 | 0.001 |
| 3/4/5/ 6 | ACCA | 30.88 | 9 | 12.01 | $0.5 \times 10^{-3}$ |
| 3/4/5/ 6 | CTTG | 12.03 | 31.04 | 8.38 | 0.004 |
| 4/5/6/ 7 | CCAA | 32.88 | 10 | 12.21 | $0.5 \times 10^{-3}$ |
| 5/6/7/ 8 | CAAG | 34.94 | 10 | 13.84 | $0.2 \times 10^{-3}$ |
| 1/2/3/ 4/5 | ATACC | 27.92 | 8.99 | 9.71 | 0.002 |
| 1/2/3/ 4/5 | GCCTT | 10 | 25.97 | 7.09 | 0.008 |
| 2/3/4/ 5/ 6 | TACCA | 30.88 | 9 | 12.01 | $0.5 \times 10^{-3}$ |
| 2/3/4/ 5/ 6 | CCTTG | 12.03 | 30.04 | 7.7 | 0.006 |
| 3/4/5/ 6/ 7 | ACCAA | 30.88 | 9 | 12.01 | $0.5 \times 10^{-3}$ |
| 4/5/6/ 7/ 8 | CCAAG | 31.94 | 10 | 11.48 | $0.7 \times 10^{-3}$ |
| 1/2/3/ 4/5/ 6 | ATACCA | 26.9 | 8 | 10.24 | 0.01 |
| 1/2/3/ 4/5/ 6 | GCCTTG | 10.04 | 22.03 | 4.48 | 0.03 |
| 2/3/4/ 5/ 6/7 | TACCAA | 30.08 | 9 | 12.01 | $0.5 \times 10^{-3}$ |
| 3/4/5/ 6/7/ 8 | ACCAAG | 29.94 | 9 | 11.26 | $0.8 \times 10^{-3}$ |
| 1/2/3/ 4/5/6/ 7 | ATACCAA | 26.9 | 8 | 10.24 | 0.001 |
| 2/3/ 4/ 5/6/7/ 8 | TACCAAG | 29.94 | 9 | 11.26 | $0.8 \times 10^{-3}$ |
| 1/2/3/ 4/5/6/ 7/8 | ATACCAAG | 26.96 | 8 | 10.29 | 0.001 |

Allele variation description in the table: 1 – rs10994982, 2 – rs7908445, 3 – rs7923074, 4 – rs10821936, 5 – rs10821937, 6 – rs7896246, 7 – rs10821938 and 8 – rs7089424

By unifying both models in one, statistically signidicant haplotypes were the more rarely seen allele in a homozygotic state. The statistically significant results are shown in table 3.7.

Table 3.7.

**_ARID5B_ gene 3rd introne allele variation risk haplotypes, connecting both study models – case control and family association models**

| SNP | Haplotype | RR (95% CI) | p value |
|---|---|---|---|
| 3/4/5 | aacccc | 3,08 (1.05 – 8.9) | 0.04 |
| 3/ 4/6 | aaccaa | 6.04 (2.14 – 16.9) | $0.8 \times 10^{-3}$ |
| 3/ 4/7 | aaccaa | 4.43 (1.63 – 11.7) | 0.005 |
| 3/ 4/8 | aaccgg | 3.29 (1.23 – 9.13) | 0.004 |
| 3/ 5/ 6 | aaccaa | 6.06 (2.09 – 17.2) | $0.6 \times 10^{-3}$ |
| 5/ 6/ 7 | ccaaaa | 7.31 (2.64 – 19.6) | $0.2 \times 10^{-3}$ |
| 5/ 6/ 8 | ccaagg | 8.67 (2.43 – 30.2) | 0.001 |
| 5/ 6/1 | ccaaaa | 7.71 (2.84 – 20.2) | $0.2 \times 10^{-3}$ |
| 6/ 7/ 8 | aaaagg | 7.36 (2.72 – 19.4) | $0.2 \times 10^{-3}$ |
| 3/ 6/ 7 | aaaaaa | 6.94 (2.61 – 18.6) | $< 0.1 \times 10^{-3}$ |
| 4/ 6/ 7 | ccaaaa | 7.03 (2.64 – 18.5) | $< 0.1 \times 10^{-3}$ |
| 5/ 6/ 7 | ccaaaa | 7.32 (2.68 – 20) | $0.2 \times 10^{-3}$ |
| 7/ 8/1 | aaggaa | 5.77 (2.31 – 14.5) | $0.6 \times 10^{-3}$ |
| 3/ 7/ 8 | aaaagg | 4.43 (1.7 – 11.5) | 0.002 |
| 5/ 7/ 8 | ccaagg | 4.88 (1.82 – 13.1) | 0.001 |
| 6/ 7/ 8 | aaaagg | 7.26 (2.7 – 19.4) | $< 0.1 \times 10^{-3}$ |
| 3/ 8/1 | aaggaa | 4.79 (1.8 – 12.3) | 0.001 |
| 4/ 8/1 | ccggaa | 5.2 (1.98 – 13.7) | 0.002 |
| 5/ 8/1 | ccggaa | 4.67 (1.68 – 12.7) | 0.003 |
| 6/ 8/1 | aaggaa | 7.03 (2.65 – 18.6) | $< 0.1 \times 10^{-3}$ |
| 3/6/6 | aaaagg | 6.05 (2.17 – 16.7) | 0.001 |
| 3/5/6 | aaccgg | 3.52 (1.23 – 10.1) | 0.02 |
| 3/4/ 5/ 6 | aaccccaa | 5.1 (1.7 – 14.8) | 0.005 |
| 3/ 4/ 5/7 | aaccccaa | 4.21 (1.51 – 11.7) | 0.006 |
| 3/ 4/ 5/8 | aaccccgg | 3.19 (1.06 – 9.6) | 0.04 |
| 3/ 4/ 6/ 7 | aaccaaaa | 6.8 (2.45 – 18.6) | $0.4 \times 10^{-3}$ |
| 3/ 4/ 6/ 8 | aaccaagg | 5.04 (1.74 – 14.2) | 0.003 |
| 3/ 4/ 7/ 8 | aaccaagg | 4.25 (1.59 – 11.2) | 0.003 |
| 3/ 5/ 6/ 8 | aaccaagg | 4.49 (1.44 – 13.6) | 0.009 |
| 4/ 5/ 6/ 8 | ccccaagg | 7.04 (1.88 – 25.5) | 0.006 |
| 4/ 5/ 7/ 8 | ccccaagg | 4.22 (1.55 – 11.8) | 0.004 |
| 5/ 6/ 7/ 8 | ccaaaagg | 5.81 (1.95 – 16.6) | 0.002 |
| 3/ 4/ 5/ 6/ 7 | aaccccaaaa | 5.81 (2.05 -17.1) | $0.6 \times 10^{-3}$ |
| 3/ 4/ 5/ 6/ 8 | aaccccaagg | 4.62 (1.49 – 12.4) | 0.006 |
| 3/ 4/ 5/ 7/8 | aaccccaagg | 4.29 (1.49 – 12.4) | 0.007 |
| 3/ 4/ 6/ 7/8 | aaccaaaagg | 5.87 (2.09 – 16.2) | 0.001 |

Allele variation description in the table: 1 – rs10994982, 2 – rs7908445, 3 – rs7923074, 4 – rs10821936, 5 – rs10821937, 6 – rs7896246, 7 – rs10821938 and 8 – rs7089424

Haplotype, which was statistically significantly linked to the risk of developing leukaemia is rs7923074/rs10821936/rs10821937/rs7896246/ rs10821938/rs7089424 – AACCCCAAAAGG RR 5,43 (95%CI 1.84 – 16.3 p value = 0.002). Graphical picture can be seen in the figure 3.4.

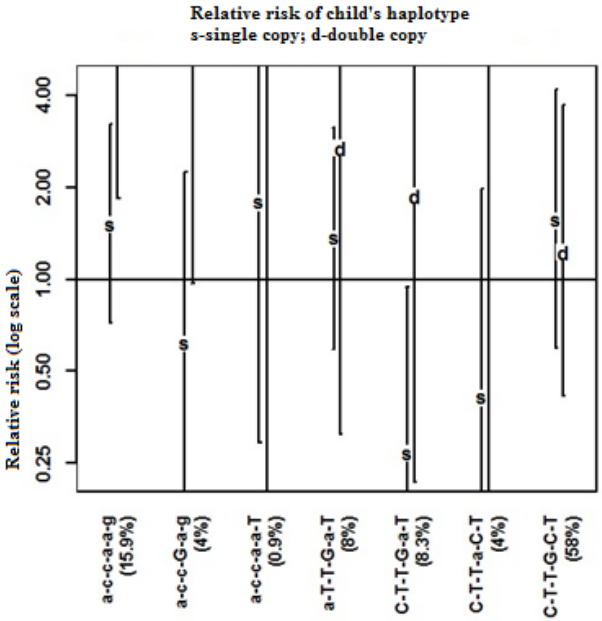


Fig. 3.4. ***ARID5B* genr allele variation risk allele haplogroups rom six allele variations, analysed by hybridmethod**

Allele variations are depicted as follows: 1– rs7923074, 2 – rs10821936, 3 – rs10821937, 4 – rs7896246, 5 – rs10821938 and 6 – rs7089424

By analysing the allele variations localized in the gene *ARID5B* they were not statistically significantly connected to the higher risk leukaemia development, nor they had a statistically significant connection to the age up to 19 years, in all cases p value > 0.05.

Analying patients' gentotypes, which ad hyperploydia (which has a better prognosis), in the comparison to patients without hyperploidia

haplotypes, connected to it were identified. Statistically significant results are depicted in table 3.8.

Table 3.8.

***ARID5B* gene allele variation haplotype analysis in patients with hyperploydia in comparison to patients without hyperploidya**

| Included allele variation haplotypes | Haplotype | Frequency in patients with hyperploidya | Frequency in patients without hyperploidya | $X^2$ | p value |
|---|---|---|---|---|---|
| 3/4//5 | ATT | 0.02 | 0.13 | 4.84 | 0.03 |
| 6/7/8 | GAT | 0.06 | 0.18 | 4.06 | 0.04 |
| 2/3/4/5 | TATT | 0.02 | 0.12 | 4.19 | 0.04 |
| 3/4/5/6 | ATTG | 0.02 | 0.13 | 4.8 | 0.028 |
| 5/6/7/8 | TGAT | 0.06 | 0.18 | 4.11 | 0.04 |
| 2/3/4/5/6 | TATTG | 0.02 | 0.12 | 4.15 | 0.04 |
| 3/4/5/6/7 | ATTGA | 0.02 | 0.14 | 4.89 | 0.026 |
| 1/2/3/4/5/6 | ATATTG | 0.02 | 0.12 | 4.14 | 0.04 |
| 2/3/4/5/6/7 | TATTGA | 0.02 | 0.12 | 4.2 | 0.04 |
| 3/4/5/6/7/8 | ATTGAT | 0.02 | 0.14 | 4.95 | 0.02 |
| 1/2/3/4/5/6/7 | ATATTGA | 0.02 | 0.12 | 4.24 | 0.04 |
| 2/3/4/5/6/7/8 | TATTGAT | 0.02 | 0.12 | 4.3 | 0.038 |
| 1/2/3/4/5/6/7/8 | ATATTGAT | 0.02 | 0.13 | 4.35 | 0.037 |

Allele variation description in the table: 1 – rs10994982, 2 – rs7908445, 3 – rs7923074, 4 – rs10821936, 5 – rs10821937, 6 – rs7896246, 7 – rs10821938 and 8 – rs7089424

## 3.3. Analysis of genes' *CEBPE, IKZF1* and *CDKN2A* allelic variations

In CEBPE gene allelic variant rs2239633 was analyzed, IKZF1 gene – rs4132601, and in gene CDKN2A – rs3731217, all three analyzed allelic variants corresponded to Hardy-Weinberg equilibrium. A statistically significant association with an increased risk of acute leukemia in this study were identified in none of the allelic variants.

In case the polymorphism rs4132601 homozygous state – GG, the child had a higher risk of developing leukemia, RR 2.77 95% CI 1.01 to 7.7, p value = 0.046.

Combining the data obtained in this study on the allelic variants rs3731217, rs2239633, rs4132601, rs10821936 and rs10994982, potential risk of combination of genotypes of ALL-risk was identified : TTCCTTCCAA, RR 9.38 (95% CI 1.56 to 58.7), p value = 0.014 and TTTTTTTTAA, 40.8 RR (95% CI 2.18 to 827) p value = 0.011.

### 3.4. *PAX5* gene allelic variation determination

When we analysed variation of s NM_001280547.1: c.547G> A; in seventy seven individuals enrolled in the study, it was not detected in any of them.

### 3.5. Genes' *IL15* allelic variations rs10519612, rs10519613 and rs17007695 analysis

When analyzed population's SNP rs10519612, rs10519613 and rs17007695, it met Hardy-Weinberg equilibrium. Analyzing each allelic variation of the possible association with risk of developing ALL, statistically significant results were found.

Before haplotype analysis we determined linkage disequilibrium (LD) between analyzed allelic variants. Between the SNP rs10519612 and rs10519613 was almost full imbalanced relationship because r2 = 0.97.

Haplotype analysis were combined in case-control and family study models. With ALL the risk statistically significantly associated haplotypes are shown in table 3.9.

Table 3.9.

**Gene's *IL15* allelic variation haplotype analaysis using hybridemthod**

| Haplotype included SNP | Haplotype | RR 95% CI | p value | Possible signficiance of haplotype |
|---|---|---|---|---|
| rs10519613/rs17007695 | AC | 5,28 (1,06 – 26) | 0,04 | risk |
| rs10519613/rs17007695 | CT | 0,21 (0,06 – 0,8) | 0,02 | protective |
| rs10519613/rs17007695 | CCTT | 0,22 (0,06 – 0,88) | 0,03 | protective |
| rs10519612/rs17007695 | AT | 0,17 (0,05 – 0,65) | 0,009 | protective |
| rs10519612/rs17007695 | AATT | 0,17 (0,04 – 0,65) | 0,009 | protective |
| rs10519612/rs105196123 / rs17007695 | ACT | 0,15 (0,04 – 0,58) | 0,006 | protective |
| rs10519612/rs105196123 / rs17007695 | AATTCC | 0,13 (0,03 – 0,52) | 0,005 | protective |

## 3.6 Results of gene allelic variations of genes involved in xenobiotic metabolism

### 3.6.1. Genes' *MDR1* allelic variations rs1045642 and rs2032582 analysis

Allelic variants rs1045642 and rs2032582 met Hardy-Weinberg equilibrium. Allelic variant rs2032582 G allele in a heterozygous state was associated with a reduced risk of developing leukemia, but the statistical credibility was lost in the homozygous state (case-control study pattern – RR 0.29; 95% CI 0.09 to 0.91; p = 0.03 ; the joint pattern – RR 0.3; 95% CI 0.1 to 0.95; p = 0.04). By contrast, homozygous condition G allele was associated with an earlier age of onset of leukemia (RR 0.13; 95% CI 0.02 to 0.18; p value = 0.03).

Allelic variant rs2032582 homozygous form of A allele was not identified. Heterozygous form of A allele was statistically significantly associated with an increased risk of developing leukemia (case-control study pattern - RR 3.5; 95% CI 1.26 to 9.51; p = 0.01; joint model - RR 3.61; 95 % CI 1.32 to 9.43; p = 0.01).

By analyzing the maternal haplotype, rs2032582 had a protective role of T allele in a homozygous state, RR 0.05; 95% CI 0.003 to 0.77; p = 0.03. By contrast, the SNP rs1045642 C allele was statistically significantly associated with a higher risk of developing leukemia in childhood (heterozygous state RR 3.09; 95% CI 1.2 to 7.79; p = 0.02; homozygous condition RR 4.49; 95% CI 1.24 to 15.6; p = 0.02).

When analyzing haplotypes, haplotype rs1045642 / rs2032582 - TA was statistically significantly associated with a higher risk of developing leukemia, RR 9.51 (95% CI 1.29 to 70.7), p value = 0.03. Protective effect on the risk of leukemia in childhood was in the maternal haplotype rs1045642 / rs2032582 TT, if both alleles are homozygous state, RR 0.09 (95% CI 0.01 to 0.85), p value = 0.035.

### 3.6.2. *MTHFR* gene allelic variation rs1801131 and rs1801133 analysis

Both analyzed allelic variants corresponded to Hardy-Weinberg equilibrium. Allelic variants had no statistically reliable relationship with the risk of developing ALL. Maternal haplotype rs1801131 / rs1801133 AACC had a protective role in the development of child' s leukemia, RR 0.13 95%CI (0.03–0.6) p value = $0.9 \times 10^{-3}$.

### 3.6.3. *NQO1* gene allelic variation rs1800566 analysis

NQO1 gene allelic variation rs1800566 met Hardy-Weinberg equilibrium and was not statistically significantly associated with risk of developing leukemia.

### 3.6.4. Deletions in genes *GSTT1* and *GSTM1*

GSTT1 gene deletion was identified in the homozygous form of 28.6% individuals ALL group, while in the control group this deletion was identified in 18.2%, but the results were not statistically significantly different, p value

> 0.05. Gene GSTM1 homozygous deletion was identified in the 55.8% subjects of ALL and 47.9% subject control group, respectively, these results were not statistically significantly different, p value> 0.05.

## 3.7. *IKZF1* Ikaros gene sequencing and protein expression analysis

Four of the sixty-eight samples (5.88%) protein was nit expressed. Expression of proteins of less than fifty percent was seen in fifteen subjects, or 22% percent. Average protein expression levels in cells was 67.65% (standard deviation 30.7). The expression level of breakdown can be seen in figure 3.5.

Analyzing Ikaros expression level depending on proband's age, statistically significant differences have not been found.

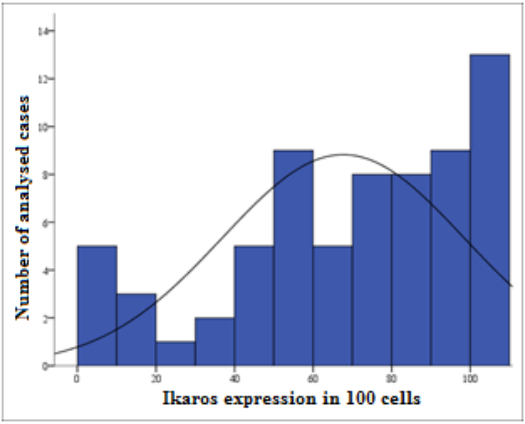


Fig. 3.5. **Ikaros protein expression in bone marrow cells**

Evaluating the protein expression level of bone marrow cells and acute leukemia relapse or exitus letalis risk statistically significant relationship was not found, however, it is noted that the blood samples were selected from 2009 onwards.

In addition to assessing the intensity of expression, samples were divided into three groups: the first included six samples with weak segmental intensity, the average expression of Ikaros in this group was 32% (standard deviation (SD) − 17); the second group included twenty-nine samples with a medium (moderate) segmental intensity with an average protein expression in 64% (SD − 17) and the third group included twenty-nine samples were identified which recognizes a total expression, with an average protein expression 91% (SD − 17). Four patients of protein expression levels were zero, so they were not included in the breakdown by group effort.

Compared all three groups with each other, a statistically significant difference was found between the first and the third group of individuals − with a weak expression intensity was lower expression levels compared with individuals who had high expression of $p < 0.1 \times 10^{-3}$ Statistical differences are represented in figure 3.6.

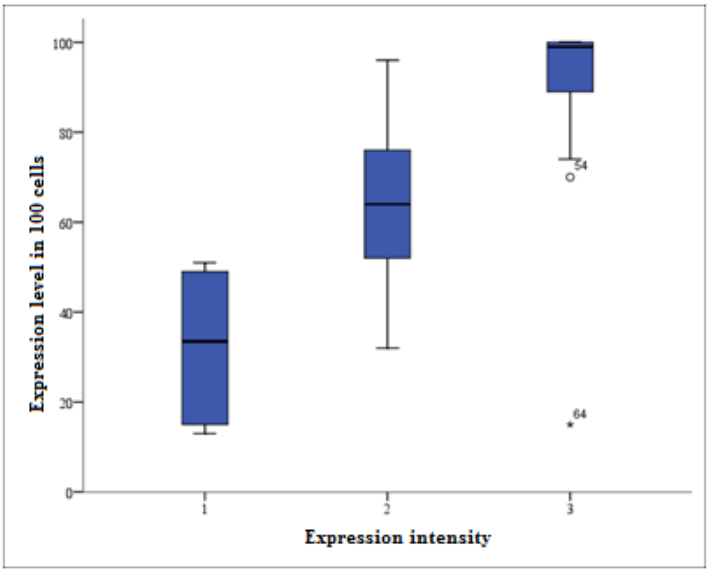


Fig. 3.6. **Ikaros protein expression intensity comparison to epxpression level**

To assess the Ikaros protein deficit also in non-somatic cells, individuals who had the expression levels of less than thirty percent in bone marrow samples and individuals whose bone marrow samples we did not have access to, immunohistochemical analysis of peripheral blood samples was performed - in all cases growth in the expression was observed of at least 30%, but in one case − by as much as 98%. In the samples,without data from bone marrow samples, expression levels in peripheral blood was greater than fifty percent.

Although there is no known reference of normal protein expression in bone marrow and peripheral blood, we assume that the expression level of relative growth in all samples indicate that the expression affecting changes have taken place in somatic cells

In gene sequencing IKZF1 second exon (the first protein coding exon) and the third exon, any sequence changes have not been found, these two exons are mandatory in all protein isoforms. Analyzing the sequences were also examined one hundred to two hundred surrounding intron nucleotides adjacent sequences both 5'direction, and 3 'direction. Analyzing the first and the second exon, allelic variants were identified.

Sequencing the fourth, fifth, sixth and seventh exon, exon, the genetic variation was not identified. In analysis of the adjacent intron nucleotides eight different allelic variants were identified based on single nucleotide point replacement: rs74412507 (MAF A = 5.6%) rs113125091 (MAF T = 6.7%), rs7789106 (MAF C = 3, 9%), rs113962761 (MAF T = 16.88), rs56278999 (MAF T = 27.27%), rs199614380 (MAF T = 2.6%), rs150440917 (MAF A = 1.3%) and rs4132098 (MAF A = 5.6%) and one allelic variant, where it has been thymine insertion - Indel rs72334180.

To assess the allelic variants of possible clinical importance were analyzed their possible involvement in the processivity using *Human Splicing finder* (*Desmet et al.*, 2009). Two of the allelic variants were detected possibly related to converting the processivity. The first is the allelic variant rs7789106,

the C allele case intron Silencer site is interrupted, the results obtained by analyzing the two detection algorithms: 1) intron identity elements can be identified introns (*C Zhang et al.*, 2008); 2) regulatory sequences in exon splicing (Goren et al., 2006). The second allelic variant, which might be involved modified processivity is rs199614380 the T allele arises jaands donor site, and this finding is confirmed by two different splaisa site analysis algorithms - jaands donor site (*Desmet et al.*, 2009; *Eng et al.*, 2004).

Analyzing rs199614380 and rs7789106, rarest allele frequency was less than five percent, and according to the literature, rare single nucleotide allelic variants of statistical power (importance) is not sufficient, especially if the number of samples is limited (*Gorlov et al.*, 2008).

Therefore these allelic variants frequency in the control group were not analyzed.

Thymine insertion ninety three nucleotides before the start of the sixth exon database Ensemble (http://www.ensembl.org/index.html) found with reference number rs72334180, whose frequency is not known.

The research group found that insertion 24.68% or 19 individuals from 77 in accordance with the potential splaisa site identifier program *Human Splicing finder* (*Desmet et al.*, 2009) consensus probability is, that this place can help processivity branching (branch) pandect is 52.82%, in the control group the incidence was 16%, a statistically significant difference between the study and control groups was not found (OR 1,72; 95%CI 0,77 – 3,86, p value = 0,18).

Analyzing gene IKZF1 eighth exon were identified two variants of synonyms. The first of these was rs61731355 NM_006060.5: c.1002C> A NP_006051.1: p.Pro334 =.

Patients with ALL, this option was identified thirty one allele A, whose frequency was 20.13%. In the control group seventy-seven individualswere analyzed by age and sex according to individuals in which A allele was

identified in seventeen alleles − 11.04%. A allele frequency of the study group were statistically significantly different from the control group (OR 2.03; 95%CI 1.03 – 4.05, p value = 0.04).

In the database *Human Splicing Finder* (*Desmet et al.*, 2009) available information shows, that in case of A allele potential exon splicing enhancers (ESE) is interrupted as a result potentially the last intron "excision" can be affected (*Liu et al.*, 2000; *C Zhang et al.*, 2008). In ESS rupture detection Cartegni and colleagues developed matrix proteins SF2 / ASF was used, consensus value changes from 76.41 to 86.43, threshold 72.98.

As the second option synonymous rs61731356 NM_006060.5 was analyzed: c.1176C> T NP_006051.1: p.Asn392 =. The research group identified T allele in seven alleles (4.54%). In the control group it was identified in nine alleles. T allele frequency of the study and the control group was not statistically significantly different. Analyzing its possible effects on the processivity, it was found that if the cytosine is replaced by thymine, turns a potential acceptor site -CCCGAGCAACAGCT CV 79.6 to cccgagcaatagCT CV 71.92 ΔCV 9.65% if jaandais acceptor site is active, in such a case 328 nucleotides may come off exon.

Taking into account the importance of protein Ikaros lymphocyte differentiation, gene IKZF1 exon deletion analysis was performed in only patient who developed acute ALL at the age under 1 year. With MLPA method no deletions were found.

# 4. DISCUSSION

Acute leukemia is the most common disease of children (*Tharnprisan et al.*, 2013). Its etiology is still not fully explored, it is considered as a multifactorial disease in its development genetic predisposition and environmental factors interaction plays a great role (*Y Yan et al.*, 2014). In this study, for the first time in Latvian population single nucleotide allelic variants were analyzed previously described in association with acute pre-B cell leukemiaand its increase the risk.

The study included two groups of allelic variants; the first of them analyzed allelic variants previously described extensive genome association studies, in addition to including allelic variants of the gene ARID5B third intron because it described as a "hot pandect" in relation to the risk of developing ALL (*Gutierrez-Camino et al.*, 2013). The second group analyzed allelic variants associated with the metabolism of xenobiotics, this group also analyzed the potential impact of maternal genotype on the risk of developing ALL, as the literature data, taking into account that the peak incidence was observed between the ages of two and five years, speculates, that environmental factors, in utero exposure and abnormalities could have a role in etiopathogenesis (*Nousome et al.*, 2013).

The study included three IL15 gene localized allelic variants, previously described only in connection with the risk of developing leukemia in adulthood (*D Lin et al.*, 2010). In addition, we analyzed one genetic variant in gene PAX5 potentially associated with the risk of developing leukemia in families (*Shah et al.*, 2013).

In the light of Ikaros protein role in lymphocyte differentiation and pathogenesis of leukemia (*Heizmann et al.*, 2013), in this study for the first time we tried to assess the genetic variant gene IKZF1 importance in non-

somatic cells, as well as to determine the protein expression levels of both leukaemic cells and leukocytes at the moment of remission.

The study included patients from 2005 onwards, when it became possible to differentiate B-cell leukemia T cell leukemia, given that the B and T lymphocytes is slightly different differentiation pathway and the genes involved (*Kalia et al.*, 2006), this study included only patients with pre-B cell leukemia. However, the sample collection was launched in 2009, whenit already nine previously diagnosed patients have died, which reduced the number of patients included, and could affect the reliability of the results, when the recurrence and mortality relationship with allelic variants were analysed.

The biggest limitation of the study was the limited number of patients, although only seven patients (or their parents) refused to participate in the study, representing 8:33% of the survey during the life of existing and without bone marrow transplant, thus, the study included 77 individuals. Mainly, patients (or their parents) refused to participate if the leukemia was diagnosedseveral years ago and the reason for the refusal was primarily emotional desire to no longer be associated with leukemia diagnosis. Despite the fact that they were only 8.33% of the total population, these patients had a long event-free period, and this fact could affect the reliability of the results related to relapse and mortality in association with genetic variations. However, the study not included in the proportion of patients who had died and who had long-term relapse, thus we can make the assumption that this factor in the effect to the results of the research is irrelevant. The study included Latvian citizens, but patients are not divided down by nationality, because children often were not detectable membership of a particular nationality, because the children were from mixed families, one of the study included the children's father was Mexican, but not a single Egyptian. This factor could affect the results, in particular the analysis of allelic variants, which are characterized by different incidences between populations.

Another factor which complicates the assessment of results is that there is no available research data on the allelic variant of the incidence and possible association with risk of developing ALL Estonia and Lithuania, which are immediate neighbors.

## 4.1. Allelic variations in the 3rd introne of gene *ARID5B*

Gene ARID5B function still is not fully clear, after previous research it is known that the gene belongs to a family of transcription factors and play an important role in embryonic development, cell type-specific gene expression and cell growth regulation. As well known that the knock-out mouse develops incorrect thymus and spleen structure, as well as impaired B cell differentiation, but the precise mechanism of action is unknown. Extensive genome association study revealed that in the third intron localized allelic variants associated with an increased risk of developing leukemia, in particular hyperlpoid leukemia cases, but sequencing third and fourth exon in the region no coding genetic variant was found (*Trevino et al.*, 2009).

Later, the results replicated in different populations including our study Latvian in the pediatric population. The most analyzed polymorfism rs10821936, rs10994982 and rs7089424 was identified extensive genome association studies (*Papaemmanuil et al.*, 2009; *Trevino et al.*, 2009). Also data about rs10821938 (*Vijayakrishnan et al.*, 2010), rs7896246 and rs7923074 (*H Xu et al.*, 2012) connection with ALL can be found in publications.

According to current information, this is the first study which demonstrated anlizēta and allelic variants rs10821937 and rs7908445 association with the risk of developing acute leukemia.

Interestingly, these allelic variants analyzed in the context of gender and probandu findings are very mixed. Two studies showed that the allelic variant allele risk is often associated with girls' gender (*Gutierrez-Camino et al.*, 2013).

In contrast, in other published studies were observed risk allele frequency differences between the sexes (*Lautner-Csorba et al.*, 2013) and there are studies whose results coincide with the results of our study that the risk allele frequency more to do with the boy gender (*Healy et al.*, 2010). These results of this research show that the most likely gender does not play a major role with ARID5B localized gene allelic variant of a link with the risk of developing leukemia.

Relatively few are also studied gene ARID5B localized allelic variant connection with the age of the development of ALL, most likely it is connected directly to the fact that relatively little is known about the mechanism of action of this gene. One publication was found, stating that the risk allele is more common in children with leukemia if it develops until the age of five and less common in children who developed leukemia after ten years of age (*Evans et al.*, 2014).

In this study, in the third intron localized allelic variants were analyzed for potential relevance to the late risk of developing leukemia, it is, after ten years of age, but statistically significant correlation was not found.

The literature describes the allelic variant rs10821936 C allele association with hyperploidy (*Trevino et al.*, 2009). Our study we analyzed in isolation only the allelic variant, and such a connection could not be found. In haplotype analysis was observed that the T allele is more common in individuals who have not been observed hyperploidy which basically coincides with the published data. It should be noted that during statistical analysis, which compared the genotype's possible link with the number of chromosomes in the leukaemic cells, analysis included information on 64 individuals, cytogenetic analysis was not available in thirteen individuals, which further reduced the test group, thus resulting statistical reliability should assessed critically.

Analyzing allelic variants in some cases control the pattern and families study results did not differ significantly, however, haplotype analysis of the reliability of the results - the odds ratio and p value between the two study models differed with a higher reliability in the family study model. According to literature data family study model has greater statistical significance of rare diseases as compared to case-control model (*De et al.*, 2013), although nowadays more and more literature haplotype analysis suggests a combination of both models in order to increase the statistical power (*Wen et al.*, 2014), which was done in tis study

Second, each of the different research models included a number of individuals, because analyzing the family study model was only available in fifty full triadi, while in the case - control model seventy seven genotypes of individuals and hundreds twenty two control individual genotypes.were analyzed

When analyzing haplotypes, it was observed that haplotypes, which was the highest in relation to the statistical reliability of ALL included allelic variant rs10821936 and / or allelic variant rs7923074, respectively, this means that on the basis of the relative risk compared to control subjects with allelic variant rs7923074 risk of developing leukemia increasing by 1.37 times, but individuals with polymorfism rs10821936 - 1.16 times.

Results of the study replicating previously published results of gene ARID5B localized allelic variant association with ALL risk, but does not show the understanding of the mechanism of action of these allelic variants, that could be related to the pathogenesis of ALL.

## 4.2. *CEBPE, IKZF1* and *CDKN2A* genes' allelic variations

One of the described allelic variants, which are associated with an increased risk of developing leukemia in children is CEBPE gene allelic variant rs2239633, although GWAS research on ALL risk was weaker compared with

the genes ARID5B and IKZF1 allelic variants p = 2,88 x $10^{-7}$ (*Papaemmanuil et al.*, 2009). GWAS results were replicated in several studies and in a meta-analysis study which analyzed the eleven publications in the closest connection with the risk of ALL was Hispanic racial Caucasian individuals, the statistical results were plausible, however, the odds ratio 95% CI was 1.09 to 1, 30th Geographically closest population, which was included in this analysis were Polish. Despite the fact that the Polish colleagues' study included patient count was much higher compared to ours, they analyzed the three hundred and ninety-eight patients genotypes, but their study ALL risk was not statistically significant with allelic variant rs2239633 (*Pan et al.*, 2014; *Pastorczak et al.*, 2011).

Interesting results were obtained by analyzing gene IKZF1 localized allelic variant rs4132601 to the G allele, which is the risk allele frequency of 29.4%, which is consistent with literature data based on the 1000 genomes project in the first phase of data of the European population MAF = 31% (*Abecasis et al.*, 2012).

However, although in the genome wide association studies, and in all performed replication studies allelic variant was statistically significantly associated with an increased risk of ALL, our study did not confirm the relationship. However, it should be noted that all trials studied population was larger. Only one study carried out in Taiwan, where the number of patients were seventy nine statistically significant relationship was not found (*CY Lin et al.*, 2014).

Most likely in a small study group this is the reason the results mismatch. However, there is a small probability that also the allelic variant in conjunction with ALL an ethnic difference was observed.

Our study identified allelic variant association with maternal genotypes that have not previously been published.

Analyzing the Ikaros protein expression level in bone marrow cells connection with the level of expression and genetic variation rs4132601 genotypes was not found, this was the first study we know of protein expression levels associated with genotype.

The literature found only one study which identified dose-dependent mRNA expression level differences depending on genotype, the risk allele was associated with a lower mRNA expression (*Papaemmanuil et al.*, 2009).

The literature has found evidence of gene CDKN2A allelic variant rs3731217 connection with risk of developing acute leukemia, this allelic variant identified in the initial of the German population study, which in order to determine possible risk variants jaandus in thirty four allelic variants were replicated in large genomic research findings. Later allelic variant rs3731217 was analyzed in Spanish, Hungaric and Canadian population (*Sherborne et al.*, 2010).

In published studies, from Poland and Thailand, which seeks to replicate the data failed to prove link to the risk of ALL (*Pastorczak et al.*, 2011; *Vijayakrishnan et al.*, 2010), just as in our study. It is likely that because this failure to identify allelic variants in the first GWAS studies, the association with the risk of ALL is not very close.

## 4.3. *PAX5* gene genetic variation

Isolated cases of ALL is also described as monogenic pathology, such as Li Fraumenn syndrome. Children with ALL is described to have TP53 gene mutation heterozygous state in non- somatic cells (*Felix, Nau, et al.*, 1992).

Currently, in families with acute leukemia gene mutation p.Gly183Ser PAX5 is found, which is inherited autosomally dominantly with incomplete penetrance, to develop acute leukemia need loss of heterozigoty is needed. Mostly in these individuals the second allele of the ninth chromosome formed

iso-chromosome., whose formation leads to loss of the short arm of the ninth chromosome (*Shah et al.*, 2013).

Patients who were diagnosed a long time ago, did not have a cytogenetic examination detailed enough and, above all, then the presence of the Philadelphia chromosomewas determined. In only of the patients who have undergone detailed cytogenetic examinations has had the ninth chromosome short arm deletion, but unfortunately the patient's parents refused to participate in the study.

Despite of this and taking into account the incomplete penetrance gene expression, it was concluded to check p.Gly183Ser presence all probands. Because parents' and researchers' main interest is focused directly on the possible monogenic inheritance ALL risks, because non-somatic cell mutations in this case, if necessary, is possible to use prenatal diagnostics.

None of the individuals had this genetic variant, despite the fact that a family history of a number of families had cases of leukemia, however, there were not specified specify the subtype,e and one child had a family history of neuroblastoma, from which her brother died and one other individual mother had died because of fatal leukemia, its type is not precisely known.

## 4.4. *IL15* gene allelic variants rs10519612, rs10519613 and rs17007695

The literature available on an increasing number of publications relating to allelic variants rs10519612, rs10519613 and rs17007695 and ALL risk in adulthood, one of newest published studies has been carried out in Egypt, where the analysis of the genetic variation was statistically significant relationship with all three allelic variants of the B-cell leukemia directly allelic variant rs17007695 genotypes CT and CC (*Aly et al.*, 2015).

IL15 in the pediatric population has been studied in the context of the MRD, which identified the relationship with allelic variant rs10519612,

rs10519613 and rs17007695, and minimal residual disease. In our study, MRD on the 33rd day were only in two patients, so the statistical analysis was not performed, because these data would be for information only.

The studies about allelic variants rs10519612 and rs10519613 are discussed separately, but in our study the allelic variants was almost in complete relation r2 = 0.97. When analyzing haplotypes genetic variation, haplotypes of the most frequently occurring alleles ALL development was aprotective role in the adult population, there the risk allele was less common, so that research results do not condradict with those publsidhed previously (*D Lin et al.*, 2010). In order to judge about the possible role of interleukin etiology of acute leukemia, more extensive studies are required, that analyze genotypes children.

## 4.5. Allelic variations in genes, that are involved in the xenobiotic metabolism

### 4.5.1. *MDR1* gene allelic variants rs1045642 and rs2032582

Despite extensive research of the genome identified allelic variants, the etiology of acute leukemia is still not entirely clear and still is believed that proteins involved in the metabolism of xenobiotics could play a role in the development of acute leukemia (*Nousome et al.*, 2013). There are hypothesis found in the publications that genetic variation in P-glycoprotein can bind with different susceptibility to environmental carcinogens and as a result of reduced enzyme activity it leads to increased risk of tumor development (*Semsei et al.*, 2008).

Analyzing the MDR1 gene allelic variant rs1045642 in this study, we identified allelic variant associated with an increased risk of leukemia, which is in line with newest meta-analysis study. Meta-analysis study analyzed a total of nine studies that included a total of 1,462 acute leukemia patients and

1,522 control individuals and identified that the allelic variant rs1045642 T allelein a homozygous conditionwas associated with an increased risk of developing leukemia only Asians, but not Caucasians (*Yue et al.*, 2015). These results could relate to the allele frequency differences between populations (*Leal-Ugarte et al.*, 2008).

Analyzing tri-allelic non-synonymous genetic variant rs2032582, the G allele was identified as potentially protective allele, but taking into account that the statistical credibility was lost in a homozygous state, these results should be assessed critically. Reliability of the results support the fact that the G allele homozygous conditionwas associated with an earlier age of onset of leukemia, which generally associated with a better prognosis. G allele as a protective potential was identified in a meta-analysis study, but overall this meta-analysis study analyzed acute myeloid leukemia patients and acute lymphoblastic leukemia patients differentiating subtypes and statistical reliability was based mainly on the p-values and a number of cases odds ratio and confidence interval was not assessed, so in this case this data must be evaluated critically. These results, in turn, could relate to the allele frequency differences between population (*Y Yan et al.*, 2014).

Our study anlizējot allelic variant rs2032582 A allele was not identified in any of individuals in a homozygous state, but its heterozygous state was identified as a potential risk allele. In the iterature A allele significance is analyzed only in individual trials found, where no statistically significant association with the risk of developing leukemia and A allele was seen, andalso no potential risk haplotypes were identified (*Semsei et al.*, 2008; *Urayama et al.*, 2007).

Combining identified genotypes allelic variants rs1045642 and rs2032582 in Latvian children population with a history of ALL, A allele allelic variant rs2032582, T allele and rs1045642 allelic variant forms a risk haplotype.

Statistically analyzing the study of families of model data identified maternal genotype relationship with risk of developing ALL, the TT haplotype decreases a child's risk of developing ALL, this is the first study that analyzes the maternal genotype relationship with the child's risk of developing ALL. However, given the origins of leukemia development can be already in utero, xenobiotics metabolising protein activity of the mother may affect the first stages of leukemia development in utero.

### 4.5.2. *MTHFR* gene allelic variants rs1801131 and rs1801133

Despite the fact that in recent years there have been a lot of studies aboutthe MTHFR gene allelic variants rs1801131 and rs1801133 and their possible link with the risk of developing ALL, the results are still contradicting. Although most of the study cohort is several times bigger compared to this study, however, also studies with small cohorts, e.g. one which included ninety-eight patients, and contrary to this study, individuals with genotypes (AC + CC) allelic variant rs1801131 had 1,1 times higher risk of developing leukemia compared to individuals whose genotype is AA (*X Li et al.*, 2014).

Certain studies have indicated ethnic differences in allelic variation, which could affect the various research result contradiction (*Giovannetti et al.*, 2008). However, the meta-analysis carried out more research points to the allelic variant rs1801133 association with the risk of developing ALL Caucasians but not in Asians (*Jiang et al.*, 2013).

Unfortunately, very little is known about the gene MTHFR allelic variants and ALL links within Eastern Europe, the publication of Slovenia analyzed allelic variant of the relationship with risk of relapse, but not the primary risk of developing leukemia (*Karas Kuzelicki et al.*, 2009).

Data from the literature on maternal genotype importance of genes encoding enzymes folate metabolism, but theu also contradict (*Lupo et al.*, 2012) and publications are relatively few. However, our study shows that if

the mother has MTHFR gene encoded enzyme activity un-altered, the child has a lower risk of developing acute leukemia compared with mothers with reduced enzyme activity.

### 4.5.3. *GSTT1* and *GSTM1* gene deletions

The study did not find a statistically significant association with the risk of developing acute leukemia and deletions glutathione S transferase genes mi and theta genes. This deletion analysis is based on a method that does not distinguish deletions' heterozygous variant form of the norm. Diagnostic methods to analyze the constraints did not allow the transfer of alleles / inheritance of family study model, as well as to assess the allele frequencies, we believe that it could have affected the overall results with respect to GSTT1 and GSTM1 deletion importance of developing ALL.

Analyzing the literature data meta-analysis of studies of each of these genes points to ethnic differences that deletion as more important than others in Asians than in Caucasians (*Moulik et al.*, 2014; *ZH Tang et al.*, 2014; *LY Xu & Cao*, 2014).

### 4.5.4. *NQO1* allelic variant rs1800566

Interesting are the results of the analysis of allelic variant rs1800566, which is localized gene NQO1, in the case TT genotype of the amino acid proline is replaced by serine, and in accordance with literature data enzyme loses its activity (*Larson et al.*, 1999). In the study group, this variant in a homozygous state was identified only one individual corresponding to 1.29%, while the control group of five individuals 4%, the results are not statistically significantly different.

Meta-analysis study shows that the T allele homozygous condition is associated with an increased risk of developing leukemia in children

(*C Li et al.*, 2014), however, the result is a mismatch, and they to be assessed critically as analyzed populations are geographically distant from the localized Latvian population, nearest of them are in Europe, Italy and United Kingdom, as well as extensive studies analyzed patients directly with early development of MLL positive ALL, i.e. up to 18 months of age (*Infante-Rivard et al.*, 2007), of course, the small number of patients should be takein into account.

## 4.6. *IKZF1* gene sequence and protein expression level

Ikaros protein is one of the key roles in differentiation of white blood cells in virtually all development stages. Lack of Ikaros stops the differentiation of pre-B cells (*Yoshida et al.*, 2014). Studies in mice have demonstrated Ikaros protein deficit is connected with the development of leukemia (*Masuda*, 2011). Analyzing the changes in somatic leukaemic cells, approximately 20−30% of cases finds mutated gene IKZF1 (*Dupuis et al.*, 2013). For example, the gene PAX5 ,which sequence changes are specific in leukaemic cells, have also been observed in non-mutated somatic cells (*Shah et al.*, 2013).

Also, it has been demonstrated thatintron allelic variant rs4132601 is associatedn with the risk of developing ALL, but the exact mechanism is not known (*Gorniak et al.*, 2014).

By analogy with the gene PAX5, this study is analyzed gene IKZF1 full coding exon sequence in seventy seven probands, no subject had been identified to have nucleotide replacement, which resulted in an amino acid replacement, which could be directly related to the risk of developing leukemia. Also, the only patient who was diagnosed with leukemia under one year of age had gene deletions IKZF1. So far, there are no available publications about IKZF1 gene mutation and gene sequence changes in the importance of children's age. During this study time there has been a publication about a child, born in 33. gestational week, with congenital pancytopenia, including

deep B-cell lymphopenia and NK deficit and unchanged T lymphocyte count. The child died on the 87$^{th}$ day of life 40 days after the bone marrow transplant from multi-organ disfunction The child was diagnosed de novo point mutation in the fifth exon (*Goldman et al.*, 2012).

Of course, based of one publication is not possible to draw conclusions, but also the absence of other publications may raise the hypothesis that the mutations localized in gene IKZF1 coding part have early clinical symptoms have been, and they are so heavy that children die before development of leukemia. In favor of this hypothesis speaks the fact that no child was identified to have a mutation in this gene as well as the analysis of protein expression in children showed tha children with bone marrow transplant had reduced protein expression levels, it renewed at the time of remission in the peripheral blood. Analyzing polimrofism rs61731355in bioinformatics tool *Human Splicing Finder*, whose incidence was statistically significant differences between the study and control groups, played a role in the ESE breaking and jaanda ESE making. Other allelic variant rs61731356 was not statistically significantly different in incidence between the study and control groups and analyzing the potential impact of bioinformatics tool showed that indicate the potential acceptor site impact on the processivity was borderline, because the CV was 79.6, but splice site from CV 80 is considered a strong splice site option and ΔCV 9.65%, which in turn ranging from 10% counts significant. However, in order to clarify this assumption would require mRNA studies.

This study also looks at one indel, which until now was of unknown incidence, but this was not statistically significantly different between the study and control groups, which are probably not clinically important.

The literature has previously described that the gene IKZF1 are highly conservative, the fact that our study did not identify mutations in this gene, presumably is also evidence that the gene is highly conservative.

# 5. CONCLUSIONS

1. Set up a study group of the seventy-seven individuals and from fifty of them, or 64.94% was available to both biological parents' genetic material. We found that most patients with ALL children were among children born in 2004 b.
2. Analyzing the patients by age, leukocyte count at the time of diagnosis, the minimal residual disease and the presence of the Philadelphia chromosome, we found that high-risk group had twenty-seven patients. We also detected, that analyzed allelic variant of localized MDR1 gene, rs2032582 G allele was associated with the development of leukemia up to ten years of age, associated with a better prognosis.
3. A statistically significant association with all of the eight analyzed gene intron 3 ARID5 allelic variants and ALL was found, as well as potential risk haplotypes were identified. For the first time an analysis of rs10821937 and rs7908445 association with the risk of developing acute leukemia was performed.
4. CEBPE, CDKN2A and IKZF1 genes allelic variants rs2239633, rs3731217 and rs4132601, did not show a statistically significant relationship. By combining the data collected on the allelic variants rs3731217, rs2239633, rs4132601, rs10821936 and rs10994982, the potential risk ALL-risk increasing combination of genotypes were identified.
5. MDR1 gene allelic variants were analysed for the risk and protective haplotypes, as well as maternal genotype effect on the progression of leukemia In the child was identified. MTHFR gene polimorfia was identified with maternal genotype effect on the risk of developing ALL. For IL15 gene allelic variants protective haplotypes were identified. Genes NQO1 and PAX5 allelic variants, as well as the genes GSTM1 and GSTT1

deletions did not have a statistically significant association with the risk of developing leukemia.

6. When the gene IKZF1 was full sequenced, pathogenic mutations were not found, two allelic variants − rs199614380 and rs7789106 were identified; one INDEL − rs72334180 and in the eighth exon localized versions of two synonyms − rs61731355 and rs61731356, which are possibly involved in splicing. Allelic variants didn not have a significant relationship with the protein expression level, none of the individuals were found to have an absolute Ikaros protein deficit in peripheral blood, which may indicate a hereditary protein deficiency.

# 6. PUBLICATIONS

**Kreile M**, Rots D, Piekuse L, Cebura E, Grutupa M, Kovalova Z, Lace B. Lack of Association between Polymorphisms in Genes MTHFR and MDR1 with Risk of Childhood Acute Lymphoblastic Leukemia. Asian Pac J Cancer Prev. 2014;15(22):9707-11

**Kreile M**, Piekuse L, Rots D, Shteinberga Z, Kovalova Zh, Lace B. Analysis of possible genetic risk factors contributing to development of childhood acute lymphoblastic leukaemia in Latvian population. Arch Med Sci.2016; 12 (3) in press

**Kreile M.,** Piekuse L., Kovaļova Ž., Cebura E., Medne G., Grūtupa M. Akūtas pre-B šunu limfoblastu leikozes bērnu vecumā aatīstības riks saistība ar alēliskie variantiem *MDR1* un *IKZF1* gēnā. Riga Stradins university, Scientific preceedings 2013, 391- 396 lpp

# 7. REFERENCES


1. Abecasis G. R., Auton A., Brooks L. D., et al. An integrated map of genetic variation from 1,092 human genomes // Nature, 2012; 491: 56-65.
2. Agostini C., Trentin L., Sancetta R., et al. Interleukin-15 triggers activation and growth of the CD8 T-cell pool in extravascular tissues of patients with acquired immunodeficiency syndrome // Blood, 1997; 90: 1115-23.
3. Alpman A., Ozkinay F., Tekgul H., et al. Multidrug resistance 1 (MDR1) gene polymorphisms in childhood drug-resistant epilepsy // J Child Neurol, 2010; 25: 1485-90.
4. Altschul S. F., Madden T. L., Schaffer A. A., et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs // Nucleic Acids Res, 1997; 25: 3389-402.
5. Aly R. M., Taalab M. M.,Ghazy H. F. Influence of interleukin-15 polymorphism on the survival of adult patients with acute lymphoblastic leukemia in Egypt // Leuk Lymphoma, 2015; 56: 151-6.
6. Ayaz G., Batar B., Kanigur G., et al. The association of MDR1 C3435T and G2677T/A polymorphisms with plasma platelet-activating factor levels and coronary artery disease risk in Turkish population // Gene, 2013; 527: 301-5.
7. Belson M., Kingsley B.,Holmes A. Risk factors for acute leukemia in children: a review // Environ Health Perspect, 2007; 115: 138-45.
8. Cordero A. M., Crider K. S., Rogers L. M., et al. Optimal serum and red blood cell folate concentrations in women of reproductive age for prevention of neural tube defects: world health organization guidelines // MMWR Morb Mortal Wkly Rep, 2015; 64: 421-3.
9. De G., Yip W. K., Ionita-Laza I.,Laird N. Rare variant analysis for family-based design // PLoS One, 2013; 8: e48495.
10. Desmet F. O., Hamroun D., Lalande M., et al. Human Splicing Finder: an online bioinformatics tool to predict splicing signals // Nucleic Acids Res, 2009; 37: e67.
11. Dunna N. R., Vure S., Sailaja K., et al. Deletion of GSTM1 and T1 genes as a risk factor for development of acute leukemia // Asian Pac J Cancer Prev, 2013; 14: 2221-4.
12. Dupuis A., Gaub M. P., Legrain M., et al. Biclonal and biallelic deletions occur in 20% of B-ALL cases with IKZF1 mutations // Leukemia, 2013; 27: 503-7.
13. Eng L., Coutinho G., Nahas S., et al. Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: maximum entropy estimates of splice junction strengths // Hum Mutat, 2004; 23: 67-76.
14. Evans T. J., Milne E., Anderson D., et al. Confirmation of childhood acute lymphoblastic leukemia variants, ARID5B and IKZF1, and interaction with parental environmental exposures // PLoS One, 2014; 9: e110255.
15. Fehniger T. A.,Caligiuri M. A. Interleukin 15: biology and relevance to human disease // Blood, 2001; 97: 14-32.
16. Felix C. A., D'Amico D., Mitsudomi T., et al. Absence of hereditary p53 mutations in 10 familial leukemia pedigrees // J Clin Invest, 1992; 90: 653-8.

17. Felix C. A., Nau M. M., Takahashi T., et al. Hereditary and acquired p53 gene mutations in childhood acute lymphoblastic leukemia // J Clin Invest, 1992; 89: 640-7.
18. Giovannetti E., Ugrasena D. G., Supriyadi E., et al. Methylenetetrahydrofolate reductase (MTHFR) C677T and thymidylate synthase promoter (TSER) polymorphisms in Indonesian children with and without leukemia // Leuk Res, 2008; 32: 19-24.
19. Goldman F. D., Gurel Z., Al-Zubeidi D., et al. Congenital pancytopenia and absence of B lymphocytes in a neonate with a mutation in the Ikaros gene // Pediatr Blood Cancer, 2012; 58: 591-7.
20. Goren A., Ram O., Amit M., et al. Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers // Mol Cell, 2006; 22: 769-81.
21. Gorlov I. P., Gorlova O. Y., Sunyaev S. R., et al. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms // Am J Hum Genet, 2008; 82: 100-12.
22. Gorniak P., Pastorczak A., Zalewska-Szewczyk B., et al. Polymorphism in IKZF1 gene affects age at onset of childhood acute lymphoblastic leukemia // Leuk Lymphoma, 2014;
23. Gutierrez-Camino A., Lopez-Lopez E., Martin-Guerrero I., et al. Intron 3 of the ARID5B gene: a hot spot for acute lymphoblastic leukemia susceptibility // J Cancer Res Clin Oncol, 2013; 139: 1879-86.
24. Hanson N. Q., Aras O., Yang F.,Tsai M. Y. C677T and A1298C polymorphisms of the methylenetetrahydrofolate reductase gene: incidence and effect of combined genotypes on plasma fasting and post-methionine load homocysteine in vascular disease // Clin Chem, 2001; 47: 661-6.
25. Healy J., Richer C., Bourgey M., et al. Replication analysis confirms the association of ARID5B with childhood B-cell acute lymphoblastic leukemia // Haematologica, 2010; 95: 1608-11.
26. Heizmann B., Kastner P.,Chan S. Ikaros is absolutely required for pre-B cell differentiation by attenuating IL-7 signals // J Exp Med, 2013; 210: 2823-32.
27. Hunger S. P., Loh M. L., Whitlock J. A., et al. Children's Oncology Group's 2013 blueprint for research: acute lymphoblastic leukemia // Pediatr Blood Cancer, 2013; 60: 957-63.
28. Inaba H., Greaves M.,Mullighan C. G. Acute lymphoblastic leukaemia // Lancet, 2013; 381: 1943-55.
29. Infante-Rivard C., Vermunt J. K.,Weinberg C. R. Excess transmission of the NAD(P)H:quinone oxidoreductase 1 (NQO1) C609T polymorphism in families of children with acute lymphoblastic leukemia // Am J Epidemiol, 2007; 165: 1248-54.
30. Jain M., Pandey P., Tiwary N. K.,Jain S. MTHFR C677T polymorphism is associated with hyperlipidemia in women with polycystic ovary syndrome // J Hum Reprod Sci, 2012; 5: 52-6.
31. Jiang Y., Hou J., Zhang Q., et al. The MTHFR C677T polymorphism and risk of acute lymphoblastic leukemia: an updated meta-analysis based on 37 case-control studies // Asian Pac J Cancer Prev, 2013; 14: 6357-62.

32. John S. W., Weitzner G., Rozen R.,Scriver C. R. A rapid procedure for extracting genomic DNA from leukocytes // Nucleic Acids Res, 1991; 19: 408.
33. Jugessur A., Shi M., Gjessing H. K., et al. Genetic determinants of facial clefting: analysis of 357 candidate genes using two national cleft studies from Scandinavia // PLoS One, 2009; 4: e5385.
34. Kalia V., Sarkar S., Gourley T. S., et al. Differentiation of memory B and T cells // Curr Opin Immunol, 2006; 18: 255-64.
35. Karas Kuzelicki N., Milek M., Jazbec J.,Mlinaric-Rascan I. 5,10-Methylenetetrahydrofolate reductase (MTHFR) low activity genotypes reduce the risk of relapse-related acute lymphoblastic leukemia (ALL) // Leuk Res, 2009; 33: 1344-8.
36. Kim Y. O., Kim M. K., Woo Y. J., et al. Single nucleotide polymorphisms in the multidrug resistance 1 gene in Korean epileptics // Seizure, 2006; 15: 67-72.
37. Kimura Y., Selmi C., Leung P. S., et al. Genetic polymorphisms influencing xenobiotic metabolism and transport in patients with primary biliary cirrhosis // Hepatology, 2005; 41: 55-63.
38. Kondo S., Sturgis E. M., Li F., et al. GSTM1 and GSTT1 null polymorphisms and risk of salivary gland carcinoma // Int J Clin Exp Med, 2009; 2: 68-75.
39. Larson R. A., Wang Y., Banerjee M., et al. Prevalence of the inactivating 609C-->T polymorphism in the NAD(P)H:quinone oxidoreductase (NQO1) gene in patients with primary and therapy-related myeloid leukemia // Blood, 1999; 94: 803-7.
40. Lautner-Csorba O., Gezsi A., Erdelyi D. J., 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: e69843.
41. Leal-Ugarte E., Gutierrez-Angulo M., Macias-Gomez N. M., et al. MDR1 C3435T polymorphism in Mexican children with acute lymphoblastic leukemia and in healthy individuals // Hum Biol, 2008; 80: 449-55.
42. Li C.,Zhou Y. Association between NQO1 C609T polymorphism and acute lymphoblastic leukemia risk: evidence from an updated meta-analysis based on 17 case-control studies // J Cancer Res Clin Oncol, 2014; 140: 873-81.
43. Li X., Liao Q., Zhang S.,Chen M. Association of methylenetetrahytrofolate reductase (MTHFR) C677T and A1298C polymorphisms with the susceptibility of childhood acute lymphoblastic leukaemia (ALL) in Chinese population // Eur J Med Res, 2014; 19: 5.
44. Lin C. Y., Li M. J., Chang J. G., et al. High-resolution melting analyses for genetic variants in ARID5B and IKZF1 with childhood acute lymphoblastic leukemia susceptibility loci in Taiwan // Blood Cells Mol Dis, 2014; 52: 140-5.
45. Lin D., Liu C., Xue M., et al. The role of interleukin-15 polymorphisms in adult acute lymphoblastic leukemia // PLoS One, 2010; 5: e13626.
46. Liu H. X., Chew S. L., Cartegni L., et al. Exonic splicing enhancer motif recognized by human SC35 under splicing conditions // Mol Cell Biol, 2000; 20: 1063-71.
47. Llaudo I., Colom H., Gimenez-Bonafe P., et al. Do drug transporter (ABCB1) SNPs and P-glycoprotein function influence cyclosporine and macrolides exposure in renal transplant patients? Results of the pharmacogenomic substudy within the symphony study // Transpl Int, 2013; 26: 177-86.

48. Lupo P. J., Nousome D., Kamdar K. Y., et al. A case-parent triad assessment of folate metabolic genes and the risk of childhood acute lymphoblastic leukemia // Cancer Causes Control, 2012; 23: 1797-803.
49. Masuda S. More on NK-cell and B-cell deficiency with a thymic mass // N Engl J Med, 2011; 364: 1979-80; author reply 80-1.
50. Mejia-Arangure J. M., Fajardo-Gutierrez A., Flores-Aguilar H., et al. Environmental factors contributing to the development of childhood leukemia in children with Down's syndrome // Leukemia, 2003; 17: 1905-7.
51. Milne E., Greenop K. R., Metayer C., et al. Fetal growth and childhood acute lymphoblastic leukemia: findings from the childhood leukemia international consortium // Int J Cancer, 2013; 133: 2968-79.
52. Misra V., Grondin A., Klamut H. J.,Rauth A. M. Assessment of the relationship between genotypic status of a DT-diaphorase point mutation and enzymatic activity // Br J Cancer, 2000; 83: 998-1002.
53. Moulik N. R., Parveen F., Kumar A.,Agrawal S. Glutathione-S-transferase polymorphism and acute lymphoblastic leukemia (ALL) in north Indian children: a case-control study and meta-analysis // J Hum Genet, 2014; 59: 529-35.
54. Mullighan C. G. The molecular genetic makeup of acute lymphoblastic leukemia // Hematology Am Soc Hematol Educ Program, 2012; 2012: 389-96.
55. Mullighan C. G., Goorha S., Radtke I., et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia // Nature, 2007; 446: 758-64.
56. Nousome D., Lupo P. J., Okcu M. F.,Scheurer M. E. Maternal and offspring xenobiotic metabolism haplotypes and the risk of childhood acute lymphoblastic leukemia // Leuk Res, 2013; 37: 531-5.
57. Pan Y., Chen H., Liang H., et al. Meta-analysis of the association between CCAAT/enhancer binding protein-epsilon polymorphism and the risk of childhood acute lymphoblastic leukemia // Int J Clin Exp Med, 2014; 7: 5553-7.
58. Papaemmanuil E., Hosking F. J., Vijayakrishnan J., et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia // Nat Genet, 2009; 41: 1006-10.
59. Pastorczak A., Gorniak P., Sherborne A., et al. Role of 657del5 NBN mutation and 7p12.2 (IKZF1), 9p21 (CDKN2A), 10q21.2 (ARID5B) and 14q11.2 (CEBPE) variation and risk of childhood ALL in the Polish population // Leuk Res, 2011; 35: 1534-6.
60. Payne K. J.,Dovat S. Ikaros and tumor suppression in acute lymphoblastic leukemia // Crit Rev Oncog, 2011; 16: 3-12.
61. Pui C. H., Robison L. L.,Look A. T. Acute lymphoblastic leukaemia // Lancet, 2008; 371: 1030-43.
62. Purcell S., Neale B., Todd-Brown K., et al. PLINK: a tool set for whole-genome association and population-based linkage analyses // Am J Hum Genet, 2007; 81: 559-75.
63. Redaelli A., Laskin B. L., Stephens J. M., et al. A systematic literature review of the clinical and epidemiological burden of acute lymphoblastic leukaemia (ALL) // Eur J Cancer Care (Engl), 2005; 14: 53-62.

64. Rosner B. Fundamentals of biostatistics // 6th ed.Published: Thomson-Brooks/Cole, 2006. -
65. Safarinejad M. R., Shafiei N.,Safarinejad S. Methylenetetrahydrofolate reductase (MTHFR) gene C677T, A1298C and G1793A polymorphisms: association with risk for clear cell renal cell carcinoma and tumour behaviour in men // Clin Oncol (R Coll Radiol), 2012; 24: 269-81.
66. Seghatoleslam A., Monabati A., Bozorg-Ghalati F., et al. Expression of UBE2Q2, a putative member of the ubiquitin-conjugating enzyme family in pediatric acute lymphoblastic leukemia // Arch Iran Med, 2012; 15: 352-5.
67. Semsei A. F., Erdelyi D. J., Ungvari I., et al. Association of some rare haplotypes and genotype combinations in the MDR1 gene with childhood acute lymphoblastic leukaemia // Leuk Res, 2008; 32: 1214-20.
68. Shah S., Schrader K. A., Waanders E., et al. A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia // Nat Genet, 2013; 45: 1226-31.
69. Sherborne A. L., Hosking F. J., Prasad R. B., et al. Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk // Nat Genet, 2010; 42: 492-4.
70. Siegel R., Ma J., Zou Z.,Jemal A. Cancer statistics, 2014 // CA Cancer J Clin, 2014; 64: 9-29.
71. Sterjev Z., Trencevska G. K., Cvetkovska E., et al. The association of C3435T single-nucleotide polymorphism, Pgp-glycoprotein gene expression levels and carbamazepine maintenance dose in patients with epilepsy // Neuropsychiatr Dis Treat, 2012; 8: 191-6.
72. Tang Q., Li J., Zhang S., et al. GSTM1 and GSTT1 null polymorphisms and childhood acute leukemia risk: evidence from 26 case-control studies // PLoS One, 2013; 8: e78810.
73. Tang Z. H., Zhang C., Cheng P., et al. Glutathione-S-transferase polymorphisms (GSTM1, GSTT1 and GSTP1) and acute leukemia risk in Asians: a meta-analysis // Asian Pac J Cancer Prev, 2014; 15: 2075-81.
74. Tharnprisan P., Khiewyoo J., Sripraya P.,Wiangnon S. Relapse-free rate with childhood acute lymphoblastic leukemia treated under the thai national protocol // Asian Pac J Cancer Prev, 2013; 14: 1127-30.
75. Trevino L. R., Yang W., French D., et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia // Nat Genet, 2009; 41: 1001-5.
76. Urayama K. Y., Chokkalingam A. P., Metayer C., et al. SNP association mapping across the extended major histocompatibility complex and risk of B-cell precursor acute lymphoblastic leukemia in children // PLoS One, 2013; 8: e72557.
77. Urayama K. Y., Wiencke J. K., Buffler P. A., et al. MDR1 gene variants, indoor insecticide exposure, and the risk of childhood acute lymphoblastic leukemia // Cancer Epidemiol Biomarkers Prev, 2007; 16: 1172-7.
78. Vijayakrishnan J., Sherborne A. L., Sawangpanich R., et al. Variation at 7p12.2 and 10q21.2 influences childhood acute lymphoblastic leukemia risk in the Thai population and may contribute to racial differences in leukemia incidence // Leuk Lymphoma, 2010; 51: 1870-4.

79. Wang H., Wang J., Zhao L., et al. Methylenetetrahydrofolate reductase polymorphisms and risk of acute lymphoblastic leukemia-evidence from an updated meta-analysis including 35 studies // BMC Med Genet, 2012; 13: 77.
80. Wen S. H.,Tsai M. Y. Haplotype association analysis of combining unrelated case-control and triads with consideration of population stratification // Front Genet, 2014; 5: 103.
81. Williams M. T., Yousafzai Y., Cox C., et al. Interleukin-15 enhances cellular proliferation and upregulates CNS homing molecules in pre-B acute lymphoblastic leukemia // Blood, 2014; 123: 3116-27.
82. Winandy S., Wu P.,Georgopoulos K. A dominant mutation in the Ikaros gene leads to rapid development of leukemia and lymphoma // Cell, 1995; 83: 289-99.
83. Woo J. S., Alberti M. O.,Tirado C. A. Childhood B-acute lymphoblastic leukemia: a genetic update // Exp Hematol Oncol, 2014; 3: 16.
84. Xu H., Cheng C., Devidas M., et al. ARID5B genetic polymorphisms contribute to racial disparities in the incidence and treatment outcome of childhood acute lymphoblastic leukemia // J Clin Oncol, 2012; 30: 751-7.
85. Xu L. Y.,Cao L. F. GSTT1 genetic polymorphism and susceptibility to childhood acute lymphoblastic leukemia: a meta-analysis // Tumour Biol, 2014; 35: 1433-7.
86. Yan J., Yin M., Dreyer Z. E., et al. A meta-analysis of MTHFR C677T and A1298C polymorphisms and risk of acute lymphoblastic leukemia in children // Pediatr Blood Cancer, 2012; 58: 513-8.
87. Yan Y., Liang H., Xie L., et al. Association of MDR1 G2677T polymorphism and leukemia risk: evidence from a meta-analysis // Tumour Biol, 2014; 35: 2191-7.
88. Yang S., Jin T., Su H. X., et al. The Association between NQO1 Pro187Ser Polymorphism and Bladder Cancer Susceptibility: A Meta-Analysis of 15 Studies // PLoS One, 2015; 10: e0116500.
89. Yoshida T.,Georgopoulos K. Ikaros fingers on lymphocyte differentiation // Int J Hematol, 2014; 100: 220-9.
90. Yousefian E., Kardi M. T.,Allahveisi A. Methylenetetrahydrofolate Reductase C677T and A1298C Polymorphism in Iranian Women With Idiopathic Recurrent Pregnancy Losses // Iran Red Crescent Med J, 2014; 16: e16763.
91. Yue Q., Xiong B., Chen L., et al. MDR1 C3435T polymorphism and childhood acute lymphoblastic leukemia susceptibility: an updated meta-analysis // Biomed Pharmacother, 2015; 69: 76-81.
92. Zhang C., Li W. H., Krainer A. R.,Zhang M. Q. RNA landscape of evolution for optimal exon and intron discrimination // Proc Natl Acad Sci U S A, 2008; 105: 5797-802.