RESEARCH ARTICLE

Editorial Process: Submission:11/01/2017 Acceptance:12/05/2017

Association of ARID5B Genetic Variants with Risk of Childhood B Cell Precursor Acute Lymphoblastic Leukaemia in Latvia

Madara Kreile*, Dmitrijs Rots, Agnese Zarina, Linda Rautiainen, Zelma Visnevska-Preciniece, Zhanna Kovalova, Linda Gailite

Abstract

Background: Acute lymphoblastic leukaemia (ALL) is the most common malignancy in childhood. Despite numerous investigations very little is still known about its aetiology. However, in one genome wide association study conducted to identify the possible genetic risk factors, two allelic variations rs10821936 and rs10994982 in the 3rd intron of the *ARID5B* gene were identified as possible ALL risk alleles. Association between *ARID5B* gene variants and ALL risk was also been confirmed for different ethnic groups. **Materials and Methods:** Eight genetic variants in the gene *ARID5B* were genotyped - rs10994982, rs7908445, rs7923074, rs10821936, rs10821937, rs7896246, rs10821938 and rs7089424 in 77 ALL patients in remission and in 122 age and gender matched controls; parental samples were also genotyped in 50 cases. **Results:** Six out of the eight (rs7908445, rs7923074, rs10821936, rs10821937, rs7896246 and rs7089424) analysed allelic variations were identified in the case-control analysis as statistically significant risk alleles for ALL development. In the family study and using hybrid analysis, all allelic variations were significantly associated with ALL. During the study, risk haplotype was identified rs10994982/rs7908445/rs7923074/ rs10821936/ rs10821937/rs7896246/rs10821938/rs7089424 – ATACCAAG – with a frequency in cases of 0.17 and in the control group at 0.29 (chi square = 6.69, p value = 0.009). In the family association study the same haplotype showed statistical significance (chi squared = 10.3, p value = 0.001). **Conclusions:** Results of the study replicate and extend previously published findings for *ARID5B* localized allelic variants, but do not explain the mechanism of action related to the pathogenesis of ALL.

Keywords: Childhood ALL- *ARID5B*- susceptibility- Latvia- genetic variants

Asian Pac J Cancer Prev, **19 (1)**, 91-95

Introduction

Acute leukaemia is most common malignancy in childhood (Hunger et al., 2013; Hashemi et al., 2016). Acute lymphoblastic leukaemia (ALL) is more common than acute myeloid leukaemia – Teshnizi and colleague had shown in their study that more than 83% from leukaemia patients have ALL (Teshnizi and Ayatollahi, 2015; Hosseini Teshnizi and Taghi Ayatollahi, 2017). The most frequent ALL subtype is B cell precursor leukaemia (Urayama et al., 2013).

Despite numerous studies very little is still known about the aetiology of ALL. (Nousome et al., 2013). Less than 5% of ALL cases are connected to genetic syndromes (Pui et al., 2008). Translocation between the chromosomes 12 and 21 encoding ETV6/RUNX1 – fusion gene is the most common chromosomal abnormality detected in 20-25% of childhood ALL cases (Inaba et al., 2013), but previous studies had shown that rearrangement of this fusion gene alone is insufficient to cause leukaemogenesis (Zakaria et al., 2017). Newest studies had shown that allelic variants in the microRNA coding genes have effects on the function of miRNA and contribute to cancer susceptibility, but still further research is needed to investigate miRNA role in the ALL pathogenesis (Rakmanee et al., 2017).

There have been several genome wide association studies conducted to identify the possible genetic risk factors. In one of the studies two allelic variations rs10821936 and rs10994982 were identified as possible ALL risk alleles which are localized in *ARID5B* gene's 3rd intron (Trevino et al., 2009). The association between *ARID5B* gene variants and ALL risk has also been confirmed for different ethnic groups (Gutierrez-Camino et al., 2013; Guo et al., 2014).

The aim of this study was to perform an extensive analysis of the *ARID5B* gene 3rd intron genetic variations association with ALL in Latvian children population.

Materials and Methods

Patients

Patients, diagnosed with acute B cell precursor

*Riga Stradiņš University, University Scientific Laboratory of Molecular Genetics, Riga Stradiņš University, Institute of Oncology, Children's Clinical University Hospital, Latvia. *For Correspondence:madara.kreile@rsu.lv*

lymphoblastic leukaemia in the time period of ten years in the Children's Clinical University Hospital were included in this study. Study was carried out in accordance with Helsinki declaration, prior to enrolment 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 group consisted of 77 patients DNA and in 50 cases DNA samples of both biological parents were available. Control group consisted of 122 DNA samples of individuals without leukaemia and was age and sex matched to the patient group. Parents of the control group individuals at the time of the sample collection had signed the informed consent form in which was stated, that these samples could be used without limitations, accordingly to the researchers discretion.

Genomic DNA Extraction and genotyping

Peripheral blood from the patients in complete remission and their biological parents and controls was used for DNA extraction. DNA isolation was performed using standard phenol/chloroform method (John et al., 1991).

Eight genetic variants in the gene *ARID5B* (OMIM # 608538) were genotyped – rs10994982, rs7908445, rs7923074, rs10821936, rs10821937, rs7896246, rs10821938 and rs7089424, variants were selected based on literature data SNPs which had MAF >5% and were in the closest localization to rs10821936. Genotyping of rs10821937, rs10821938 and rs10994982 was performed using polymerase chain reaction – restriction fragment length polymorphism (PCR-RFLP) analysis, primer sequences (Metabion, Germany) and restriction enzymes (Thermo Fisher, USA) used for the PCR-RFLP assay are shown in the Table 1. Genotyping of rs7908445, rs7923074, rs10821936, rs7896246 and rs7089424 was performed by Sanger sequencing reactions using Big Dye 3.0 Terminator kit according to the manufacturer protocol (Thermo Fisher, USA) primer sequences shown in Table 1.

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 association with ALL was performed, using the PLINK 1.07 software (Purcell et al., 2007), correction by sex was performed.

For case control and family study model statistical result gathering, and for increasing the statistical confidence in these results, a hybrid analysis was used, in which analysis was performed, using the R software, Haplin package, which is based on a log-linear model, and if necessary, uses – expectation–maximization algorithm for haplotype reconstruction (Gjessing and Lie, 2006; Hosseini Teshnizi and Taghi Ayatollahi, 2017).

Results

In total we analysed eight allelic variations, localized in the *ARID5B* gene's 3rd intron, whose MAF > 5%. All allelic variation genotype distribution matched Hardy-Weinberg equilibrium.

Six out of the eight analysed allelic variations in case-control analysis were identified as statistically significant risk alleles for ALL development. After correction for sex the results remained significant. Analysing allele frequency by sex, it has been found, that the risk allele was statistically significantly more frequent in boys than girls. The case-control allele association analysis is shown in the Table 2.

Evaluating the allelic variations' association with the risk of ALL development in the family study, all analysed allelic variations were statistically significantly associated to ALL development. Results are shown in Table 3.

For enforcing the statistical confidence of these results, both study models – case control and patient-parent trio studies were unified by hybridmethod (Gjessing and Lie, 2006; Jugessur et al., 2009). Results are shown in Table 4.

The possible risk haplotypes were analysed in both

Table 1. Primer Sequencies with Annealing Temperatures and Restriction Enzymes Used in PCR-RFLP Assay

Gene's allelic variation(#db SNP)	Primer sequence	Annealing temperature, ToC	Restriction enzyme
rs10821937	F 5' AAA ACA TTG CCA GGA TCT GC 3' R 5' TAA ATA GCT GCT GCC CCA TC 3'	60	Eco81I
rs10821938	F 5' TGG TGC TTT GGG GTA GTT TT 3' R 5' AAA AAT ACC CTG CCC CTT TG 3'	59	BstxI
rs10994982	F 5' CTC TTG AAC TCC TGG CCT CA 3' R 5' GCG TCA CCA AAC TCA GCT AA 3'	60	SatI
rs7908445 rs7923074 rs10821936	F 5'CAG TGA GAG CGA GAC TGC AC 3' R 5' TCC TGG CAA TGT TTT TCA CA 3'	60	-
rs7896246	F 5'GAG GCC ATT CTA GTG CGT TC 3' R 5'ACT ACC CCA AAG CAC CAT TG 3'	60	-
rs7089424	F 5' GGC CAT GAA GTC TCA CCT GT 3' R 5'TTT CAA ACC CAA AAC CAA GG 3'	60	-

SNP, single nucleotide polymorphism

Table 2. Case- Control Analysis of the Gene *ARID5B* 3rd Intron Variants

Gene's allelic variation (#db SNP)	Minor allele	MAF in patient group	MAF in control group	p value	OR (CI 95%)
rs10994982	A	0.51	0.41	0.054	1.5 (0.99-2.26)
rs7908445	T	0.43	0.32	0.022	1.63 (1.07-2.47)
rs7923074	A	0.44	0.32	0.016	1.67 (1.1-2.54)
rs10821936	C	0.34	0.24	0.027	1.65 (1.05-2.56)
rs10821937	C	0.34	0.23	0.018	1.7 (1.09-2.68)
rs7896246	A	0.34	0.23	0.018	1.7 (1.09-2.68)
rs10821938	A	0.47	0.38	0.074	1.45 (0.96-2.18)
rs7089424	G	0.34	0.23	0.018	1.7 (1.09-2.68)

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

Table 3. Family (Trio) Analysis of Variations Located in 3rd Intron of Gene *ARID5B*

Gene's allelic variations (#db SNP)	OR (CI 95%)	p value
rs7908445	2.53 (1.39-4.61)	0.002
rs7923074	2.79 (1.51-5.13)	0.0006
rs10821936	2.62 (1.38 -4.96)	0.002
rs10821937	3.18 (1.62-6.27)	0.0004
rs7896246	2.9 (1.41-5.95)	0.002
rs10821938	1.91 (1.12- 3.230	0.015
rs7089424	3 (1.56-5.77)	0.0005
rs10994982	1.88 (1.05-3.39)	0.032

OR, odds ratio; CI, confidence interval

study models – case-control and family study. The haplotype composed from analysed variations in case control group which showed statistical significance was rs10994982/rs7908445/rs7923074/ rs10821936/ rs10821937/rs7896246/rs10821938/rs7089424 – ATACCAAG – frequency in cases 0.17, in control group 0.29 (chi square = 6.69, p value = 0.009). In family association study the same haplotype showed statistical significance (chi square = 10.29, p value = 0.001).

Analysing patients' variations genotypes and haplotypes, who were hyperdiploid in comparison with patients who were not, haplotype rs10994982/ rs7908445/rs7923074/rs10821936/rs10821937/ rs7896246/rs10821938/rs7089424 – ATATTGAT – was statistically significantly more frequent in hyperdiploid group comparing with non-hyperdiploid – 0.13 and 0.02, respectively, chi square = 4.35, p value = 0.037 .

Discussion

Acute leukaemia is the most common malignant disorder in children (Tharnprisan et al., 2013). Its aetiology 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 (Yan et al., 2014). In this study, for the first time in Latvian population variants in gene *ARID5B* intron 3rd were analysed in association with acute B precursor cell leukaemia.

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 (Trevino et al., 2009). Analysed allelic variants previously described extensive genome association studies, in addition to included allelic variants of the gene *ARID5B* 3rd intron because it is described as a "hot spot" in relation to the risk of developing ALL (Gutierrez-Camino et al., 2013). Later, the results replicated in different populations including our study – Latvian paediatric population. The most frequently reported variants rs10821936, rs10994982 and rs7089424 were identified in extensive genome association studies (Papaemmanuil et al., 2009; Trevino et al., 2009). Also data about rs10821938 (Vijayakrishnan et al., 2010), rs7896246 and rs7923074 (Xu et al., 2012) association with ALL was reported.

Table 4. Hybrid Analysis of Variants Located in 3rd Intron of Gene *ARID5B*

	One risk allele in genotype		Two risk alleles in genotype	
Gene's allelic variations (#db SNP)	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)	0.0004
rs10821937	1.35 (0.76-2.35)	0.312	5.29 (2.32-11.9)	0.0002
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)	0.0002
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

According to current information, this is the first study which demonstrated allelic variants rs10821937 and rs7908445 association with the risk of developing ALL. In contrast to other published studies our study did not show statistically significant association between rs1099498, rs10821938 and ALL even after correction for sex, for example correction for sex in Yemeni children showed more marked association in females (Linabery et al., 2013; Al-Absi et al., 2017). There are studies whose results coincide with the results of our study that the risk allele frequency is higher for male sex (Healy et al., 2010).

Similar to previously published articles variants which showed strongest association with leukaemia development were rs10821936 and rs7089424 (Bhandari et al., 2016) and additionally rs10821937, which has not been described before.

The literature describes the allelic variant rs10821936 C allele association with hyperdiploidy (Trevino et al., 2009). In the haplotype analysis performed in our study was observed that the T allele is more common in individuals who have not been observed hyperdiploidy which is in concordance with the published data.

There were some limitations of our study, the sample collection was launched, when already nine previously diagnosed patients had died, which reduced the number of patients included, and could affect the reliability of the results. 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, despite the small percentage 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 did not include 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 were not divided by nationality. Children often had not detectable membership of a particular nationality, some were from mixed families, father of one child included in the study was Mexican, in the second family father was Egyptian. This factor could affect the results, in particular the analysis of allelic variants, which are characterized by different incidences among populations.

Another factor which complicates the assessment of results is the absence of available research data on the allelic variant of the incidence and possible association with risk of developing ALL in Estonia and Lithuania, which are immediate neighbours of Latvia.

Analysing allelic variants in case-control study and family-trio study results did not differ significantly. According to the literature data family study model has greater statistical significance for rare diseases in comparison to case-control model (De et al., 2013), although nowadays more and more literature suggests haplotype analysis a combination of both models in order to increase the statistical power (Wen and Tsai, 2014), which was done in this study.

Results of the study replicate previously published results of gene *ARID5B* localized allelic variants as well as previously not studied variant association with ALL risk, but do not explain the mechanism of action of these allelic variants that could be related to the pathogenesis of ALL.

References

Al-Absi B, Noor SM, Saif-Ali R, et al (2017). Association of *ARID5B* gene variants with acute lymphoblastic leukemia in Yemeni children. *Tumour Biol*, **39**, 1010428317697573.

Bhandari P, Ahmad F, Mandava S, et al (2016). Association of genetic variants in *ARID5B*, IKZF1 and CEBPE with risk of Childhood de novo B-Lineage Acute Lymphoblastic leukemia in India. *Asian Pac J Cancer Prev*, **17**, 3989-95.

De G, Yip WK, Ionita-Laza I, et al (2013). Rare variant analysis for family-based design. *PLoS One*, **8**, e48495.

Gjessing HK, Lie RT (2006). Case-parent triads: estimating single- and double-dose effects of fetal and maternal disease gene haplotypes. *Ann Hum Genet*, **70**, 382-96.

Guo LM, Xi JS, Ma Y, et al (2014). *ARID5B* gene rs10821936 polymorphism is associated with childhood acute lymphoblastic leukemia: a meta-analysis based on 39,116 subjects. *Tumour Biol*, **35**, 709-13.

Gutierrez-Camino A, Lopez-Lopez E, Martin-Guerrero I, et al (2013). Intron 3 of the *ARID5B* gene: a hot spot for acute lymphoblastic leukemia susceptibility. *J Cancer Res Clin Oncol*, **139**, 1879-86.

Hashemi M, Bahari G, Naderi M, et al (2016). Association of lnc-LAMC2-1:1 rs2147578 and CASC8 rs10505477 polymorphisms with risk of childhood acute lymphoblastic leukemia. *Asian Pac J Cancer Prev*, **17**, 4985-9.

Healy J, Richer C, Bourgey M, et al (2010). Replication analysis confirms the association of *ARID5B* with childhood B-cell acute lymphoblastic leukemia. Haematologica, 95, 1608-11.

Hosseini Teshnizi S, Taghi Ayatollahi SM (2017). Comparison of cox regression and parametric models: Application for assessment of survival of pediatric cases of acute leukemia in Southern Iran. *Asian Pac J Cancer Prev*, **18**, 981-5.

Hunger SP, Loh ML, Whitlock JA, et al (2013). Children's oncology group's 2013 blueprint for research: acute lymphoblastic leukemia. *Pediatr Blood Cancer*, **60**, 957-63.

Inaba H, Greaves M, Mullighan CG (2013). Acute lymphoblastic leukaemia. *Lancet*, **381**, 1943-55.

John SW, Weitzner G, Rozen R, et al (1991). A rapid procedure for extracting genomic DNA from leukocytes. *Nucleic Acids Res*, **19**, 408.

Jugessur A, Shi M, Gjessing HK, et al (2009). Genetic determinants of facial clefting: analysis of 357 candidate genes using two national cleft studies from Scandinavia. *PLoS One*, **4**, e5385.

Linabery AM, Blommer CN, Spector LG, et al (2013). *ARID5B* and IKZF1 variants, selected demographic factors, and childhood acute lymphoblastic leukemia: a report from the children's oncology group. *Leuk Res*, **37**, 936-42.

Nousome D, Lupo PJ, Okcu MF, et al (2013). Maternal and offspring xenobiotic metabolism haplotypes and the risk of childhood acute lymphoblastic leukemia. *Leuk Res*, **37**, 531-5.

Papaemmanuil E, Hosking FJ, Vijayakrishnan J, et al (2009). Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet*, **41**, 1006-10.

Pui CH, Robison LL, Look AT (2008). Acute lymphoblastic leukaemia. *Lancet*, **371**, 1030-43.

Purcell S, Neale B, Todd-Brown K, et al (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, **81**, 559-75.

Rakmanee S, Pakakasama S, Hongeng S, et al (2017). Increased risk of Thai childhood acute lymphoblastic leukemia with the MiR196a2 T>C polymorphism. *Asian Pac J Cancer Prev*, **18**, 1117-20.

Teshnizi SH, Ayatollahi SM (2015). A comparison of logistic regression model and artificial neural networks in predicting of student's academic failure. *Acta Inform Med*, **23**, 296-300.

Tharnprisan P, Khiewyoo J, Sripraya P, et al (2013). Relapse-free rate with childhood acute lymphoblastic leukemia treated under the thai national protocol. *Asian Pac J Cancer Prev*, **14**, 1127-30.

Trevino LR, Yang W, French D, et al (2009). Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet*, **41**, 1001-5.

Urayama KY, Chokkalingam AP, Metayer C, et al (2013). SNP association mapping across the extended major histocompatibility complex and risk of B-cell precursor acute lymphoblastic leukemia in children. *PLoS One*, **8**, e72557.

Vijayakrishnan J, Sherborne AL, Sawangpanich R, et al (2010). 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*, **51**, 1870-4.

Wen SH, Tsai MY (2014). Haplotype association analysis of combining unrelated case-control and triads with consideration of population stratification. *Front Genet*, **5**, 103.

Xu H, Cheng C, Devidas M, et al (2012). *ARID5B* genetic polymorphisms contribute to racial disparities in the incidence and treatment outcome of childhood acute lymphoblastic leukemia. *J Clin Oncol*, **30**, 751-7.

Yan Y, Liang H, Xie L, et al (2014). Association of MDR1 G2677T polymorphism and leukemia risk: evidence from a meta-analysis. *Tumour Biol*, **35**, 2191-7.

Zakaria Z, Othman N, Ismail A, et al (2017). Whole-exome sequencing of ETV6/RUNX1 in four childhood acute lymphoblastic leukaemia cases. *Asian Pac J Cancer Prev*, **18**, 1169-75.