RESEARCH ARTICLE

Association of Genetic Variants in ARID5B, IKZF1 and CEBPE with Risk of Childhood *de novo* B-Lineage Acute Lymphoblastic Leukemia in India

Prerana Bhandari[1], Firoz Ahmad[1], Swarna Mandava[2], Bibhu Ranjan Das[1]*

Abstract

Background: Childhood acute lymphoblastic leukemia (ALL) is a heterogeneous genetic disease and its etiology remains poorly understood. Recent genome wide association and replication studies have highlighted specific polymorphisms contributing to childhood ALL predispositions mostly in European populations. It is unclear if these observations generalize to other populations with a lower incidence of ALL. The current case-control study evaluated variants in ARID5B (rs7089424, rs10821936), IKZF1 (rs4132601) and CEBPE (rs2239633) genes, which appear most significantly associated with risk of developing childhood B-lineage ALL. Materials and Methods: Using TaqMan assays, genotyping was conducted for 162 de novo B-lineage ALL cases and 150 unrelated healthy controls in India. Appropriate statistical methods were applied. Results: Genotypic and allelic frequencies differed significantly between cases and controls at IKZF1-rs4132601 (p=0.039, p=0.015) and ARID5B-rs10821936 (p=0.028, p=0.026). Both rs10821936 (p=0.019; OR 0.67; 95% CI=0.47–0.94) and rs4132601 (p=0.018; OR 0.67; 95%CI 0.48-0.94) were associated with reduced disease risk. Moreover, gender-analysis revealed male-specific risk associations for rs10821936 (p=0.041 CT+CC) and rs4132601 (p=0.005 G allele). Further, ARID5B-rs7089424 and CEBPE-rs2239633 showed a trend towards decreased disease risk but without significance (p=0.073; p=0.73). Conclusions: Our findings provide the first evidence that SNPs ARID5B-rs10821936 and IKZF1-rs4132601 are associated with decreased B-lineage ALL susceptibility in Indian children. Understanding the effects of these variants in different ethnic groups is crucial as they may confer different risk of ALL within different populations.

Keywords: Childhood ALL - ARID5B - IKZF1 - CEBPE - genetic susceptibility

Asian Pac J Cancer Prev, **17 (8)**, 3989-3995

Introduction

Acute lymphoblastic leukemia (ALL) is a leading cause of death in pediatric cancers (Kreile et al., 2014). Approximately 85% of all pediatric ALL is B-lineage ALL. In developing countries, including India, cure-rates of children with cancer is inferior (~10-30%) compared to high-income countries (~80%) (Jiang et al., 2013; Yadav et al., 2014; Pongstaporn et al., 2015). Further, treatment is not optimal and etiology of ALL remains poorly understood. Early diagnosis and prevention of childhood ALL still remain major challenges in clinical practice and urge the search of novel disease markers. It is increasingly being recognised that inherited genetic factors contribute to ALL susceptibility. This was complemented by recent genome wide association studies (GWAS) on European populations that identified single nucleotide polymorphisms (SNPs) at three novel loci: 7p12.2 (IKZF1), 10q21.2 (ARID5B) and 14q11.2 (CEBPE) associated with pediatric ALL risk (Treviño et al., 2009; Papaemmanuil et al., 2009). These robust associations have been subsequently evaluated in multiple ALL studies and different ethnic populations like European-white (Prasad et al., 2010; Healy et al., 2010; Pastorczak et al., 2011; Gutierrez-Camino et al., 2013), African-American (Yang et al., 2010; Xu et al., 2012), Hispanics (Chokkalingam et al., 2013; Walsh et al., 2013) and Asian (Han et al., 2010; Vijayakrishnan et al., 2010; Wang et al., 2013; Lin et al., 2014b).

These SNPs are among the strongest cancer susceptibility variants identified through GWAS with a nearly threefold risk of disease, suggesting that inherited factors greatly contribute to childhood ALL (Fletcher and Houlston, 2010). Among these SNPs, the strongest association was provided by variant rs4132601 mapping to 3'UTR (untranslated region) of IKZF1 (Ikaros zinc finger 1) gene, a tumor suppressor and master regulator of lymphocyte development, differentiation and function (Georgopoulos et al., 1994). Somatic aberrations of IKZF1 occur in 10-15% childhood ALL cases and carry a poorer outcome than patients with unmutated IKZF1 (Mullighan et al., 2009).

*[1]Research and Development, [2]Cytogenetic Division, SRL Ltd., Goregaon (W), Mumbai, India *For correspondence: brdas@srl.in*

The second strongest association was attained with SNPs in intron 3 of ARID5B (AT-rich interactive domain5B) gene, a transcription factor essential for embryogenesis and growth regulation (Treviño et al., 2009; Papaemmanuil et al., 2009). Accruing proof suggests that ARID family members are involved in cancer-related signal pathways, extremely mutated or differentially expressed in neoplasm tissues, and act as predictive factors for disease prognosis or therapeutic outcome (Lin et al., 2014a). In concordance, previous studies highlight significant associations of ARID5B SNPs with anti-leukemic drug (methotrexate) accumulation (Treviño et al., 2009), B-hyperdiploid ALL-subtype (Papaemmanuil et al., 2009; Healy et al., 2010; Gutierrez-Camino et al., 2013), relapse-risk (Xu et al., 2012) and with childhood ALL risk in blacks, contributing to racial differences in leukemia incidence (Treviño et al., 2009; Yang et al., 2010; Xu et al., 2012). Further, variant rs2239633 provided the third strongest association and mapped to CEBPE (CCAAT/enhancer-binding protein-ε) gene, a transcription factor involved in myelopoiesis regulation (Papaemmanuil et al., 2009).

Understanding the impact of these variants in various populations is crucial as previous studies highlight that inherited genetic variations in different races are related to different risk levels of ALL (Yang et al., 2010). ALL incidence substantially differs between populations, with higher rates in European and low rates in Asian populations, which may reflect genetic and environmental heterogeneity (Vijayakrishnan et al., 2010). Moreover, majority of the studies on these risk factors confine to western populations or focus on individuals of European descent, displaying a lack of population diversity, and limited studies exist in Asia. Thus, replication is crucial to confirm their impact and to discover genetic differences among them for better prediction or outcome. At present, no study on these polymorphisms in ALL exists in India. With this background, the current case-control study evaluated variants in ARID5B, (rs7089424, rs10821936), IKZF1 (rs4132601) and CEBPE (rs2239633) genes, most significantly associated with risk of developing childhood B-lineage ALL.

Materials and Methods

Study Subjects

The present retrospective study was performed at the R&D Division, SRL Ltd., Mumbai, India and included total 312 subjects, referred from various cancer hospitals and clinics across India to our center. A structured questionnaire with detailed information pertaining to demographic factors, current and past medical history, lifestyle habits and family history of cancer for each participant was collected through perusal of their medical records and personal interviews. Written consent was obtained from patients and parents of minors, before sample collection. The study was approved by the Institutional ethics committee in accordance with the Declaration of Helsinki. Selection criteria of previous ALL-association studies involving these SNPs were followed. Thus, 150 healthy subjects, aged 10-60yrs., with no previous or concurrent malignancy, served as controls. The cases included 162 de novo B-lineage ALL patients, aged ≤18yrs, diagnosed by routine cytogenetic and/or molecular analysis at our center, as described previously (Bhandari et al., 2015). Treatment and outcome was not available and hence not analyzed. Table 1 depicts the distribution of demographic parameters between cases and controls. Blood specimens were collected in Becton Dickinson vacutainer (Franklin Lakes, NJ USA) EDTA tubes and stored at -20°C until DNA isolation.

Genomic DNA Extraction and SNP Genotyping

Genomic DNA was isolated from specimens (peripheral blood or bone marrow) using QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) and stored at -20°C until further analysis. Genotyping of SNPs (ARID5B rs10821936, rs7089424; IKZF1 rs4132601 and CEBPE rs2239633) was conducted by TaqMan allelic discrimination assay on ABI 7500Fast Real-Time-PCR system (Applied Biosystems). Supp. Table 1 provides characteristics of SNPs and TaqMan-Assays. Genotype calls were made upon visualization of allelic discrimination charts in which the clusters were identified by comparison with reference controls for each SNP. To ensure quality of genotyping, we included blind duplicates (inter- and intra-plate), negative controls per assay, randomly repeated about 10-20% samples for replication and concordance was absolute.

Statistical Analysis

Routine statistical analysis were performed using SPSS v15 (SPSS Inc., Chicago, IL) and SNPStats software (Sole et al., 2006). For each SNP, deviation of genotype distribution from Hardy-Weinberg equilibrium (HWE) was assessed by Fisher's exact test among controls. Pearson's χ^2 test was used to detect differences in genotype and allele frequencies. To determine the association with risk of B-lineage ALL, odds ratios (OR) with 95% confidence intervals (CI) were measured using logistic regression. Gender-specific associations were calculated via stratified analysis comparing male cases to male controls and female cases to female controls. Gene-dosage effect was analysed by genotype-based trend test (additive model), which reveals associations that depend additively upon the minor allele. Linkage disequilibrium (LD) and haplotype analysis were conducted. Haplotype-specific ORs were estimated using the most frequent haplotype as reference. A likelihood ratio test was performed to evaluate global haplotype association with disease status and reported as Global P value. All statistical tests were two-tailed, and $p \leq 0.05$ was considered statistically significant. GraphPad StatMate 2.0 (GraphPad Software, Inc.) was used to calculate power. At a significance level of 0.05 and a minor allele frequency (MAF) of 0.5, our study can reach >80% power to detect ORs of 1.3-1.5. With less frequent SNPs (assuming MAF of 0.2), our study can reach 55% .

Results

Alleles and Genotypes Distribution, and Risk correlation

Distribution of genotypes for each SNP was in HWE among controls (p>0.05). As shown in Table 2, among cases and controls, the genotype distribution of rs10821936 (ARID5B) and rs4132601 (IKZF1) differed significantly (p=0.028; p=0.039). Likewise risk allele frequency (RAF) for rs10821936 ('C' 0.61 *vs*. 0.52, p=0.026) and rs4132601 ('G' 0.37 *vs*. 0.28, p=0.015) showed statistically significant differences. For rs7089424 (ARID5B) and rs2239633 (CEBPE), neither genotype nor allele distribution showed significant differences (both p>0.05), although a higher proportion of rs7089424 risk allele 'G' was seen in cases (0.60 *vs*. 0.53). Furthermore, Male cases-controls at rs4132601 showed significant differences in genotype (p=0.018) and allele (p=0.005) frequency (Table 2). Figure 1 shows the observed genotype patterns for each SNP.

Table 3 encompasses the association of each SNP with childhood B-lineage ALL risk. At rs10821936, variant allele C conferred a reduced disease risk (p=0.019; OR 0.67; 95% CI=0.47-0.94). Also, the carriers of heterozygous risk genotype CT (p=0.02), homozygous risk

Table 1. Distribution of Demographic Parameters

Parameters	Total (n=312)	
	Cases (n=162)	Controls (n=150)
Gender		
Males, n (%)	99 (61)	89 (59)
Females, n (%)	63 (39)	61 (41)
M:F ratio	1.57	1.46
Hemoglobin (gm/dl) Median (range)	7.4 (1.6-15)	13.9 (8.3-17.2)
Platelets (x 10^9/L) Median (range)	40 (0.7-470)	264.5 (135-470)
WBC (x 10^9/L) Median (range)	20.3 (0.8-536)	7.3 (4.1-13.1)
Normal Karyotype, n (%)	68 (42)	-
Chromosomal abnormalities, n (%)		
Hyperdiploidy	15 (9)	-
Hypodiploidy	24 (15)	-
t(9;22)/ BCR-ABL	19 (12)	-
t(1;19)/ E2A-PBX1	16 (10)	-
t(12;21)/ ETV6-RUNX1	8 (5)	-
t(4;11)/ MLL-AF4	6 (4)	-

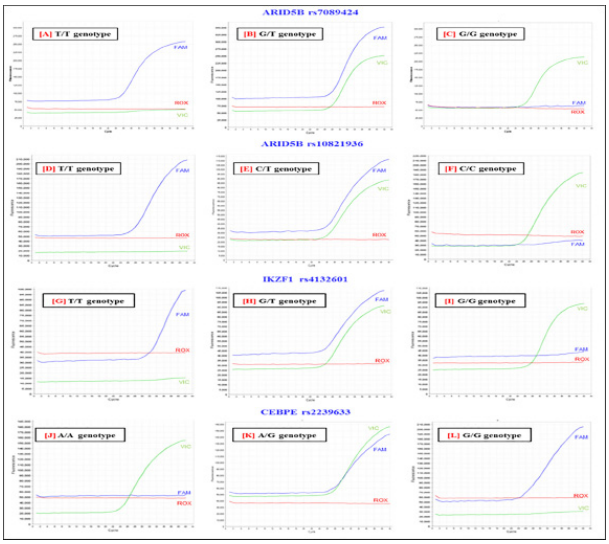

Figure 1. Genotype Signal Patterns of SNPs: showing ARID5B rs7089424 (A) T/T, (B) G/T, (C) G/G; ARID5B rs10821936 (D) T/T, (E) C/T, (F) C/C; IKZF1 rs4132601(G) T/T, (H) G/T, (I) G/G and CEBPE rs2239633 (J) A/A, (K) A/G, (L) G/G genotypes respectively. The alleles were detected by 6-carboxyfluorescein (FAM)-labeled probes and 2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein (VIC)-labeled probes

Table 2. Distribution of Genotypes and Alleles

Genotype & Allele	Total (n=312)			Males (n=188)		
Count (proportion)	Cases	Controls	P value*	Cases	Controls	P value*
ARID5B rs7089424						
TT	19 (0.12)	29 (0.19)	0.14	13 (0.13)	19 (0.21)	0.29
GT	93 (0.57)	83 (0.55)		55 (0.56)	47 (0.53)	
GG	50 (0.31)	38 (0.25)		31 (0.31)	23 (0.26)	
G[a]	193 (0.60)	159 (0.53)	0.098	117 (0.59)	93 (0.52)	0.18
T	131 (0.40)	141 (0.47)		81 (0.41)	85 (0.48)	
ARID5B rs10821936						
TT	18 (0.11)	33 (0.22)	0.028*	13 (0.13)	22 (0.25)	0.12
CT	91 (0.56)	78 (0.52)		55 (0.56)	44 (0.49)	
CC	53 (0.33)	39 (0.26)		31 (0.31)	23 (0.26)	
C[a]	197 (0.61)	156 (0.52)	0.026*	117 (0.59)	90 (0.51)	0.096
T	127 (0.39)	144 (0.48)		81 (0.41)	88 (0.49)	
IKZF1 rs4132601						
TT	69 (0.43)	78 (0.52)	0.039*	37 (0.37)	47 (0.53)	0.018*
GT	67 (0.41)	61 (0.41)		45 (0.45)	37 (0.42)	
GG	26 (0.16)	11 (0.07)		17 (0.17)	5 (0.06)	
Ga	119 (0.37)	83 (0.28)	0.015*	79 (0.40)	47 (0.26)	0.005*
T	205 (0.63)	217 (0.72)		119 (0.60)	131 (0.74)	
CEBPE rs2239633						
AA	21 (0.13)	17 (0.11)	0.57	13 (0.13)	10 (0.11)	0.29
AG	65 (0.40)	69 (0.46)		40 (0.40)	46 (0.52)	
GG	76 (0.47)	64 (0.43)		46 (0.46)	33 (0.37)	
G[a]	217 (0.67)	197 (0.66)	0.73	132 (0.67)	112 (0.63)	0.44
A	107 (0.33)	103 (0.34)		66 (0.33)	66 (0.37)	

[a] risk allele; * P≤0.05, statistically significant difference

Table 3. Correlation between SNP Genotypes and Risk of Pediatric B-lineage ALL

SNP ID	OR (95% CI)	P value	Male cases *vs* controls	
Genotype			OR (95% CI)	P value
ARID5B rs7089424				
GT *vs* TT	0.58 (0.31-1.12)	0.10	0.58 (0.26-1.31)	0.19
GG *vs* TT	0.50 (0.24-1.02)	0.055	0.51 (0.21-1.23)	0.13
GT+GG *vs* TT	0.55 (0.30-1.04)	0.062	0.56 (0.26-1.21)	0.13
G	0.73 (0.52-1.03)	0.073[a]	0.73 (0.47-1.13)	0.16[a]
ARID5B rs10821936				
CT *vs* TT	0.47 (0.24-0.89)	0.02*	0.47 (0.21-1.04)	0.061
CC *vs* TT	0.40 (0.20-0.81)	0.011*	0.44 (0.18-1.05)	0.062
CT+CC *vs* TT	0.44 (0.24-0.83)	0.009*	0.46 (0.22-0.98)	0.041*
C	0.67 (0.47-0.94)	0.019*[a]	0.69 (0.45-1.06)	0.085[a]
IKZF1 rs4132601				
GT *vs* TT	0.81 (0.50-1.29)	0.37	0.65 (0.35-1.19)	0.16
GG *vs* TT	0.37 (0.17-0.81)	0.011*	0.23 (0.08-0.69)	0.006*
GT+GG *vs* TT	0.68 (0.44-1.07)	0.096	0.53 (0.30-0.95)	0.033*
G	0.67 (0.48-0.94)	0.018*[a]	0.54 (0.35-0.85)	0.0057*[a]
CEBPE rs2239633				
AG *vs* AA	1.31 (0.64-2.70)	0.46	1.49 (0.59-3.78)	0.39
GG *vs* AA	1.04 (0.51-2.14)	0.91	0.93 (0.37-2.38)	0.88
AG+GG *vs* AA	1.17 (0.59-2.30)	0.66	1.19 (0.50-2.88)	0.69
G	0.95 (0.68-1.31)	0.73[a]	0.85 (0.55-1.30)	0.45[a]

OR, odds ratio; CI, confidence interval; * P≤0.05, statistically significant difference; [a]P value calculated by Log-additive inheritance to test for trend in association with copies of minor allele

Table 4. Distribution of SNP Haplotypes and Haplotype Risk Estimates

SNP variant ID				Haplotype frequency estimation		OR[a] (95% CI)	P value	Global P value[b]
ARID5B	ARID5B	IKZF1	CEBPE					
rs7089424	rs1081936	rs4132601	rs2239633	Cases	Controls			
Total (n=312)								
G	C	T	G	0.26	0.22	1	-	0.041*
T	T	T	G	0.17	0.26	1.70 (0.90 - 3.23)	0.11	
G	C	G	G	0.14	0.1	0.75 (0.33 - 1.68)	0.48	
G	C	T	A	0.13	0.11	1.02 (0.42 - 2.49)	0.96	
T	T	T	A	0.05	0.1	3.61 (1.02 - 12.72)	0.047*	
T	T	G	A	0.1	0.03	0.35 (0.09 - 1.42)	0.14	
G	C	G	A	0.05	0.08	1.81 (0.67 - 4.93)	0.24	
T	T	G	G	0.07	0.06	1.32 (0.49 - 3.53)	0.59	
Males (n=188)								
G	C	T	G	0.29	0.21	1	-	0.016*
T	T	T	G	0.15	0.27	2.66 (1.03 - 6.86)	0.044*	
G	C	G	G	0.13	0.1	1.15 (0.40 - 3.31)	0.8	
G	C	T	A	0.09	0.12	2.84 (0.68 - 11.78)	0.15	
T	T	T	A	0.05	0.13	4.44 (1.17 - 16.94)	0.03*	
T	T	G	A	0.12	0.04	0.33 (0.08 - 1.42)	0.14	
G	C	G	A	0.06	0.07	2.06 (0.49 - 8.71)	0.33	
T	T	G	G	0.07	0.04	1.22 (0.31 - 4.86)	0.78	

OR, odds ratio; CI, confidence interval; -, not applicable; [a]haplotype-specific ORs estimated in either overall or subgroups stratified by gender. The most common haplotype was used as reference; [b] Global P value calculated by likelihood ratio test to compare global haplotype association between cases and controls; * P≤0.05, statistically significant difference

genotype CC (p=0.011) and combined CT+CC genotype (p=0.009) achieved statistical significance (Table 3).

Association at rs7089424 and rs2239633 also showed a trend towards decreased disease risk but missed significance (p=0.073; p=0.73) (Table 3).

Regarding rs4132601, risk allele G showed decreased disease susceptibility (p=0.018; OR 0.67; 95% CI=0.48-0.94). Moreover, the carriers of homozygous risk genotype GG (p=0.011) were statistically significant (Table 3).

Using gender-based analysis, reduced disease risk was observed in males at rs10821936 (CT+CC, p=0.041; OR 0.46; 95% CI=0.22-0.98) and at rs4132601 (G allele, p=0.005; OR 0.54; 95% CI=0.35-0.85 and carriers of risk genotypes GG, p=0.006; GT+GG, p=0.033) (Table 3).

Linkage Disequilibrium Analysis and Haplotype Distributions

Haplotype analysis was performed to decipher any association between disease and individual haplotypes (Table 4). The third most frequent haplotype among cases carried the risk alleles (GCGG) whereas the non-risk alleles (TTTA) formed the complementary haplotype. We observed significant difference in the overall distribution of derived haplotypes (Global P=0.041) and an association with TTTA haplotype combination (p=0.047; OR 3.61; 95% CI=1.02-12.72) (Table 4). Similar results were seen

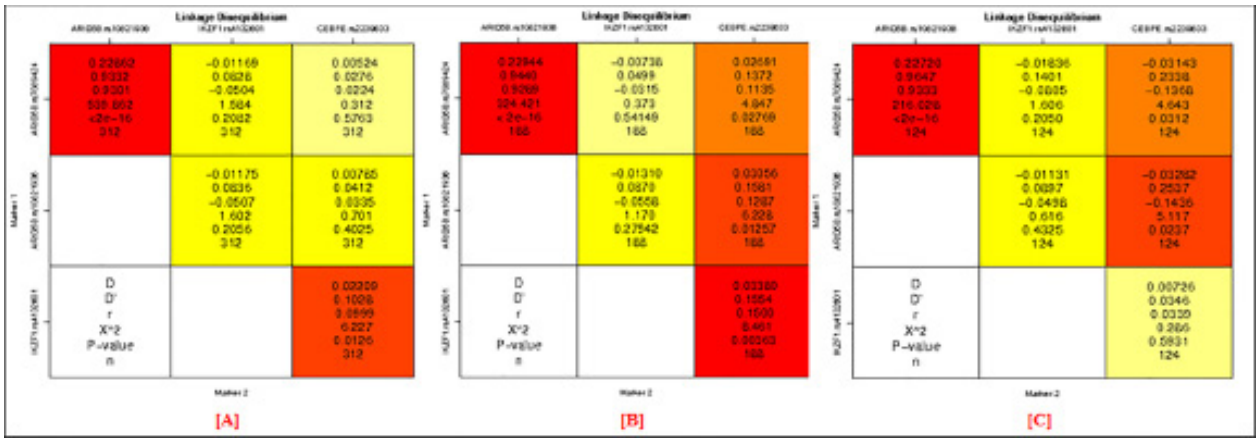

Figure 2. Linkage Disequilibrium Analysis: HeatMap in (A) Total Cohort, (B) Only Males and (C) Only Females

in males (Global P= 0.016, TTTG haplotype p=0.044; OR 2.66; 95% CI=1.03-6.86, TTTA haplotype p=0.03; OR 4.44; 95% CI=1.17-16.94) (Table 4).

Finally, LD analysis revealed strong correlations between ARID5B SNPs rs7089424-rs10821936: D'=0.93, r=0.93. Likewise, sex-specific LD patterns showed similar strong LD in males and females respectively (r^2>0.80, all p<0.0001). All other r^2 values for LD between these and other SNPs were <0.20 (Figure. 2).

Discussion

A high proportion of cancers arise by cumulative effects of common low-penetrance alleles and rare disease-causing variants that individually confer only moderate cancer risks. Thus, identifying disease-specific susceptibility (risk) and protective markers can aid in immune-genetic profiling, risk assessment and therapeutic decisions.

The majority of studies reporting strong risk associations for childhood ALL at 7p12.2, 10q21.2 and 14q11.2 loci are focused on European regions and non-European populations are undisputedly underrepresented. To the best of our knowledge, this is the first association study to investigate the frequency, distribution and association of these SNPs with childhood B-lineage ALL risk in India.

International HapMap data indicated interethnic variation in allelic distribution of SNPs examined in the current study (www.hapmap.org). Risk allele frequencies (RAF) for rs10821936, rs7089424 and rs4132601 show wide variation in different global populations. For rs10821936, 'C' allele frequency was lowest (0.16-0.2) in Kenyan, Nigerian and African ancestry populations and further varied (0.30-0.39) in other European and Asian populations. Likewise, frequency of 'G' allele for rs7089424 (0.21-0.6) and 'G' allele for rs4132601 (0.08-0.36) varied widely. Strikingly, the highest RAF of both ARID5B (0.60 'C' and 'G') and IKZF1 (0.36 'G') recorded in Gujarati Indians in USA (GIH) was similar to findings in our study (0.52, 0.53 and 0.28 respectively in controls).

In this study, rs10821936 and rs4132601 were associated with pediatric B-lineage ALL risk: risk allele carriers in both SNPs ('C' and 'G' allele) showed a decreased risk of developing disease. The contribution of these variants to ALL risk varies across different populations (Yang et al., 2010). An increased disease risk was reported, for rs4132601 in UK (Papaemmanuil et al., 2009; Prasad et al., 2010), Poland (Pastorczak et al., 2011), USA (Chokkalingam et al., 2013), France-Australia (Orsi et al., 2012) and Thailand (Vijayakrishnan et al., 2010), and for rs10821936 in various Caucasian reports (Treviño et al., 2009; Yang et al., 2010; Healy et al., 2010; Xu et al., 2012; Orsi et al., 2012; Gutierrez-Camino et al., 2013; Chokkalingam et al., 2013; Kennedy et al., 2015) and a single Asian study from China (Wang et al., 2013). Notably, significant differences in rs4132601 allelic frequency were noted between Thai and Caucasian populations (RAF ratio=0.36 for Thai/Caucasian) (Vijayakrishnan et al., 2010). Alternatively, rs4132601 showed no risk association in French-Canadian population (Healy et al., 2010) and rest available Asian studies (Han et al., 2010; Wang et al., 2013; Lin et al., 2014b). Recent meta analysis have concludedthat IKZF1-rs4132601 is associated with B-lineage ALL risk in Europeans and Hispanics, but not in Asians (Li et al., 2014; Dai et al., 2014). Such result variations may be attributed to several factors like differences in ethnicity, LD structure, allelic frequencies, environmental background interacting with variants, genetic makeup and small sample set.

We observed no evidence of risk association with ARID5B-rs7089424 and CEBPE-rs2239633 variants. Concordant with our results for rs7089424 were available Asian reports in Thailand and Taiwan (Vijayakrishnan et al., 2010; Lin et al., 2014b). Likewise, for rs2239633 no association was reported from China (Wang et al., 2013), Thailand (Vijayakrishnan et al., 2010), Poland (Pastorczak et al., 2011) and Canada (Healy et al., 2010). Conversely in several European studies, rs7089424 showed an increased risk (Treviño et al., 2009; Papaemmanuil et al., 2009; Prasad et al., 2010; Healy et al., 2010; Pastorczak et al., 2011; Ellinghaus et al., 2012; Orsi et al., 2012; Gutierrez-Camino et al., 2013; Chokkalingam et al., 2013; Walsh et al., 2013; Kennedy et al., 2015). Regarding rs2239633, few reports from UK (Papaemmanuil et al., 2009; Prasad et al., 2010) and USA (Chokkalingam et al., 2013; Walsh et al., 2013) indicated an increased disease risk, while

others showed decreased disease risk (Orsi et al., 2012; Ellinghaus et al., 2012). Strikingly, recent meta-analysis confirm that based on ethnicity, ARID5B-rs7089424 and CEBPE-rs2239633 are only associated with childhood ALL risk in Caucasians and not Asians (Zeng et al., 2014; Pan et al., 2014). This could explain the lack of association in our population. We hypothesize that such variations may reflect existence of true population-specific disease variants, but can also arise from population differences in LD pattern at these loci (differences in LD between tagging SNPs and causal variants).

Regarding gender-specific effects, in the current study ARID5B-rs10821936 and IKZF1-rs4132601 were associated with disease risk in males. This effect is not unexpected given the previous sex-specific associations observed in childhood ALL. In accordance with our results for rs10821936, Healy et al. (Healy et al., 2010) reported a male bias. Conversely, the ESCALE-GWAS on French-Australian population reported a female bias (Orsi et al., 2012), whereas Xu et al. (Xu et al., 2012) and Lautner-Csorba et al. (Lautner-Csorba et al., 2012) found no gender-specific effect.

Additionally, in our LD analysis, rs10821936, the strongest association signal observed in previous European GWAS (Treviño et al., 2009), was in LD with rs7089424 (r=0.93). Moreover, similar strong LD (r^2>0.80) was observed between rs7089424-rs10821936 in both males and females. In agreement with our findings, in another study on French-Canadian population of European descent SNPs rs10821936 and rs7089424 were highly correlated (r=0.95) and strongly associated with childhood B-lineage ALL risk (Healy et al., 2010).

In the present study, characterization of ARID5B, IKZF1 and CEBPE polymorphisms in healthy Indian individuals gives for the first time the allelic frequencies that could be used as a reference in other epidemiological studies. Given the rarity of ALL, consistency of association observed and concordance with previous published data indicates good validity and sensitivity of our study. In association studies, the true underlying genetic model is nknown. In the present study, the use of different genetic models provided strength to determine associations that would remain undetected by exclusive use of a single model.

In the current study, association with ALL-subgroups was not evaluated due to their low frequencies. Further, identifying such effects will be only realistically possible through multi-center pooled analyses, especially for countries with lower ALL incidence like India. Hence, further larger studies and thorough subgroup analyses are warranted. Few SNPs of our study did not reveal significant associations; nevertheless, this data may contribute to future meta-analysis studies. Probable explanations towards the lack of replication might include different genotype frequencies or restricted power of our study to identify loci with weaker effects or the complexity underlying ALL pathogenesis, which may differ across studies. Lastly, ethnic background may influence the risk; thus the hitherto reported risk factors might not be valid for all geographical parts of the world.

In summary, our results suggest a protective effect of ARID5B and IKZF1 variants against B-lineage ALL in Indian children. Understanding the impact of these variants in various populations is crucial as they associate with different risk levels of ALL in different races owing to genetic, ethnic and environmental heterogeneity.

References

Bhandari P, Ahmad F, Dalvi R, et al (2015). Cytogenetic Profile of De Novo B lineage Acute Lymphoblastic Leukemia: Determination of Frequency, Distribution Pattern and Identification of Rare and Novel Chromosomal Aberrations in Indian Patients. *Asian Pac J Cancer Prev*, **16**, 7219-29.

Chokkalingam AP, Hsu L-I, Metayer C, et al (2013). Genetic variants in ARID5B and CEBPE are childhood ALL susceptibility loci in Hispanics. *Cancer Causes Control*, **24**, 1789-95.

Dai Y, Tang L, Healy J, Sinnett D (2014). Contribution of Polymorphisms in IKZF1 Gene to Childhood Acute Leukemia: A Meta-Analysis of 33 Case-Control Studies. *PLoS ONE*, **9**, 113748.

Ellinghaus E, Stanulla M, Richter G, et al (2012). Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic. *Leukemia*, **26**, 902-9.

Fletcher O, Houlston RS (2010). Architecture of inherited susceptibility to common cancer. *Nat Rev Cancer*, **10**, 353-61.

Georgopoulos K, Bigby M, Wang J-H, et al (1994). The ikaros gene is required for the development of all lymphoid lineages. *Cell*, **79**, 143-56.

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.

Han S, Lee K-M, Park SK, et al (2010). Genome-wide association study of childhood acute lymphoblastic leukemia in Korea. *Leuk Res*, **34**, 1271-4.

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

Jiang Y, Hou J, Zhang Q, et al (2013). 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*, **14**, 6357-62.

Kennedy AE, Kamdar KY, Lupo PJ, et al (2015). Genetic markers in a multi-ethnic sample for childhood acute lymphoblastic leukemia risk. *Leuk Lymphoma*, **56**, 169-74.

Kreile M, Rots D, Piekuse L, et al (2014). Lack of association between polymorphisms in genes MTHFR and MDR1 with risk of childhood acute lymphoblastic leukemia. *Asian Pac J Cancer Prev*, **15**, 9707-11.

Lautner-Csorba O, Gezsi A, Semsei AF, et al (2012). Candidate gene association study in pediatric acute lymphoblastic leukemia evaluated by Bayesian network based Bayesian multilevel analysis of relevance. *BMC Med Genomics*, **5**, 1-15.

Li S, Ren L, Fan L, Wang G (2014). IKZF1 rs4132601 polymorphism and acute lymphoblastic leukemia susceptibility: a meta-analysis. Leuk Lymphoma, 56, 1-13.

Lin C, Song W, Bi X, et al (2014a). Recent advances in the ARID family: focusing on roles in human cancer. *OncoTargets Ther*, **7**, 315-24.

Lin C-Y, Li M-J, Chang J-G, et al (2014b). High-resolution melting analyses for genetic variants in ARID5B and IKZF1 with childhood acute lymphoblastic leukemia susceptibility

loci in Taiwan. *Blood Cells Mol Dis*, **52**, 140-5.
Mullighan CG, Su X, Zhang J, et al (2009). Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med*, **360**, 470-80.
Orsi L, Rudant J, Bonaventure A, et al (2012). Genetic polymorphisms and childhood acute lymphoblastic leukemia: GWAS of the ESCALE study (SFCE). *Leukemia*, **26**, 2561-4.
Pan Y, Chen H, Liang H, Wang X, Wang L (2014). Meta-analysis of the association between CCAAT/enhancer binding protein-ε polymorphism and the risk of childhood acute lymphoblastic leukemia. *Int J Clin Exp Med*, **7**, 5553-7.
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.
Pastorczak A, Gorniak P, Sherborne A, et al (2011). 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*, **35**, 1534-6.
Pongstaporn W, Pakakasama S, Chaksangchaichote P, et al (2015). MDR1 C3435T and C1236T polymorphisms: association with high-risk childhood acute lymphoblastic leukemia. *Asian Pac J Cancer Prev*, **16**, 2839-43.
Prasad RB, Hosking FJ, Vijayakrishnan J, et al (2010). Verification of the susceptibility loci on 7p12.2, 10q21.2, and 14q11.2 in precursor B-cell acute lymphoblastic leukemia of childhood. *Blood*, **115**, 1765-7.
Sole X, Guino E, Valls J, Iniesta R, Moreno V (2006). SNPStats: a web tool for the analysis of association studies. *Bioinformatics*, **22**, 1928-9.
Treviño LR, Yang W, French D, et al (2009). Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet*, **41**, 1001-5.
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.
Walsh KM, Chokkalingam AP, Hsu L-I, et al (2013). Associations between genome-wide Native American ancestry, known risk alleles and B-cell ALL risk in Hispanic children. *Leukemia*, **27**, 2416-9.
Wang Y, Chen J, Li J, et al (2013). Association of three polymorphisms in ARID5B, IKZF1and CEBPE with the risk of childhood acute lymphoblastic leukemia in a Chinese population. *Gene*, **524**, 203-7.
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.
Yadav SP, Rastogi N, Kharya G, et al (2014). Barriers to Cure for children with cancer in india and strategies to improve outcomes: a report by the indian pediatric hematology oncology group. *Pediatr Hematol Oncol*, **31**, 217-24.
Yang W, Treviño LR, Yang JJ, et al (2010). ARID5B SNP rs10821936 is associated with risk of childhood acute lymphoblastic leukemia in blacks and contributes to racial differences in leukemia incidence. *Leukemia*, **24**, 894-6.
Zeng H, Wang XB, Cui NH, et al (2014). Associations between AT-rich Interactive Domain 5B gene Polymorphisms and Risk of Childhood Acute Lymphoblastic Leukemia: a Meta-analysis. *Asian Pac J Cancer Prev*, **15**, 6211-7.