## Germline variants in *IKZF1*, *ARID5B*, and *CEBPE* as risk factors for adult-onset acute lymphoblastic leukemia: an analysis from the GMALL study group

A number of single nucleotide polymorphisms (SNPs) in the genes *IKZF1, ARID5B, CEBPE* and *CDKN2A* have been implicated in the propensity to develop acute lymphoblastic leukemia (ALL) in childhood (< 15 years of age)[1-4] but since the genetic background of adult and childhood ALL is different, it is not clear if these are also risk factors for adult-onset ALL. We analyzed seven SNPs (two in *IKZF1*: rs11978267, rs4132601; three in *ARID5B*: rs10821936, rs10994982, rs7089424; and one each in *CEBPE*: rs2239633 and *CDKN2A*: rs3731217) which were reported to confer an increased risk for childhood ALL in 322 immunologically and genetically well-characterized adult or adolescent ALL patient samples. All samples were obtained from *BCR-ABL*-negative patients at time points when they were in molecular remission, thus excluding mutations acquired during onset of disease and randomly selected from a large collection of archived samples obtained during diagnostic procedures in the framework of the German Multicenter ALL Study Group (GMALL) trial 07/2003 (*Clinicaltrials.gov identifier:00198991*). This trial has been approved by various local and central ethics committees and our study complied with the principles set forth in the Declaration of Helsinki. The median age of the patient cohort was 38 years (range 16-76 years), and 58.4% of the patients were male. Flow cytometry immunophenotyping and molecular genetic analyses for *BCR-ABL, MLL-AF4, MLL-ENL, TCF3-PBX1*, and *ETV6-RUNX1* were performed at primary diagnosis as described previously.[5,6] Assessment of minimal residual disease was made as described.[7,8] *BCR-ABL*-positive samples and mature B-ALL (sIg$^+$, or *MYC-IGH*$^+$) samples were excluded. Twenty-six patients were *MLL-AF4*-positive, 5 *MLL-ENL*-positive, 2 *TCF3-PBX1*-positive, and 2 *ETV6-RUNX1*-positive. Patient samples were selected in such a way as to match the distribution of immunophenotypes observed in the last ten years in the GMALL study group, thereby taking into consideration that *BCR-ABL*-positive patients had been excluded (Table 1).

The control group comprised a total of 1516 healthy individuals of German origin collected through 2004 at the Institute of Transfusion Medicine in Mannheim, Germany (genotyped for rs4132601, rs7089424, rs2239633, rs3731217).[4,9] Since no German control data were available for *ARID5B* rs10994982, this SNP was compared to a cohort comprising 14,311 individuals of European descent without ALL, collected by the Wellcome Trust Case Control Consortium (WTCCC); these results are only reported for completeness and are not listed in Table 2.[10] Details of the control groups have been described previously.[2,4] The SNP Genotyping Assays (Applied Biosystems, Darmstadt/Germany, Cat # 4351379) were used with the TaqMan Genotyping Master Mix (Applied Biosystems, Cat # 4371355) on a RotorGene cycler (QIAGEN, Hilden/Germany) using basically the conditions recommended by the supplier.

Some SNPs in the same gene were very closely correlated suggesting that they belonged to gene haplotypes. Linkage disequilibrium (LD) analysis revealed the following correlations: *IKZF1* rs4132601-rs11978267: D'=0.9931, $r^2$=0.9794; *ARID5B* rs7089424-rs10821936: D'=0.9731, $r^2$=0.9407, *ARID5B* rs10821936-rs10994982: D'=0.9305, $r^2$=0.5017; and *ARID5B* rs7089424-rs10994982: D'=0.9009, r2=0.4735 (see *Online Supplementary Appendix*). Therefore, rs11978267 and rs10821936 were not considered in further analysis since they added nothing new.

The observed allele frequencies are summarized in Table 2. All presumed "risk alleles" were more frequent in B-lineage ALL patients as compared to non-ALL controls while there was no difference in frequencies in T-lineage ALL. For the following genotypes, the differences were statistically significant in B-lineage ALL patients: *IKZF1* rs4132601 ('G/T', *P*=9.2E-05; 'G/G', *P*=1.4E-04), *ARID5B* rs7089424 ('G/G', *P*=2.0E-03), and *CEBPE* rs2239633 ('C/C', *P*=4.7E-02). Four genotype differences narrowly missed the significance threshold: *ARID5B* rs7089424 'G/T' (*P*=0.055), *CDKN2A* rs3731217 'T/T' (*P*=0.094), and *CEBPE* rs2239633 ('C/T', *P*=0.064). Male patients had lower overall frequencies of all risk genotypes than female patients which was largely but not entirely explained by the higher percentage of T-lineage ALL in males.[11] Females with *MLL*-negative B-lineage ALL appeared at particular high risk with ORs of 4.74 (*P*=6.75E-07; 2.567;8.768) for rs4132601 'G/G' and 1.87 (*P*=1.5E-02; 1.130;3.086) for rs4132601 'G/T', but the numbers were small (19 and 37 of a total of 84 *MLL*-negative female B-lineage ALL patients with 'G/G' and 'G/T' genotype) and, therefore, this finding needs to be validated in an independent sample set. The previous pediatric studies reported a moderate risk for childhood-onset ALL with ORs of 1.8-2.8 (*IKZF1* rs4132601),[1,4] 1.6-3.2 (*ARID5B* rs7089424),[1,4] 1.1-1.6 (*CEBPE* rs22396339),[1,4] and 1.4 (*CDKN2A* rs3731217).[3] In our adult cohort, we observed comparable ORs for rs4132601 (1.5-2.0) and rs22396339 (1.6) but the ORs for rs7089424 were lower and only significant for the bialleleic SNP (Table 2). rs10994982 showed no difference as compared to the WTCCC control group.

In summary, the ORs were similar for pediatric and adult ALL, and thus the risk conferred, appears to be lifelong and not limited to childhood. The lower ORs for *ARID5B* in adult ALL might be explained by the lower frequency of hyperdiploid B-lineage ALL in adults.[1] To address the question as to how these intronic SNPs contribute to leukemogenesis, we investigated CD34$^+$ stem cells for cryptic splice variants caused by the *IKZF1* and *ARID5B* SNPs but no splice variants were detected.

Recently, Peyrouze *et al*. reported their findings when investigating 150 adult ALL patients for almost the same six SNPs in the context of the GRAALL study group.[12] No association of any of the investigated SNPs was found but the study had some limitations, most notably a relatively small patient number and an apparently heterogeneous (including *BCR-ABL*$^+$) patient population and control group. Since these genetic risk factors confer only a moderate risk for adult- or childhood-onset ALL, efforts

**Table 1.** Immunophenotypes of the investigated patients. *BCR-ABL*-positive patients (25% of the total) were excluded.

| Immunophenotype | N. of investigated samples (%) |
|---|---|
| Common ALL (CD10$^+$/CD19$^+$/cyIg-/sIg$^-$) | 131 (40.7%) |
| Pre B ALL (CD10$^+$/CD19$^+$/cyIg$^+$/sIg$^-$) | 43 (13.4%) |
| Pro B ALL (CD10$^-$/CD19$^+$/sIg-) | 53 (16.5%) |
| Thymic T ALL (CD1a$^+$/CD2+/CD7+) | 48 (14.9%) |
| Mature T ALL (CD1a$^-$/sCD3$^+$($^-$)/CD7$^+$) | 23 (7.1%) |
| Early T ALL (CD1a$^-$/sCD3-/CD7$^+$/CD2$^-$($^+$)) | 24 (7.5%) |

Table 2. SNP genotypes in patients and controls. Absolute and relative frequencies of different SNP genotypes in patients (entire cohort, B-lineage patients, T-lineage patients) and controls, with corresponding odds ratios and *P* values. The following genotypes were risk factors for B-lineage ALL: rs4132601 'G/T' and 'G/G', rs7089424 'G/G', rs2239633 'C/C'. None of the genotypes was a risk factor for T-lineage ALL.

| SNP | Genotypes and risk allele | All patients | | B lineage patients | | T lineage patients | | Control group | | OR [95%CI] (all patients/ B lineage/ T lineage) | *P* (all patients/ B lineage/ T lineage) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N. | Freq | N. | Freq | N. | Freq | N. | Freq | | |
| rs4132601 (*IKZF1*) | | | | | | | | | | | |
| | T/T | 137 | 0.425 | 86 | 0.379 | 51 | 0.537 | 811 | 0.540 | | |
| | G/T | 151 | 0.469 | 111 | 0.489 | 40 | 0.421 | 574 | 0.382 | 1.6 [1.2;2.0]/ 1.8 [1.3;2.5]/ 1.1 [0.7;1.7] | 6.6E-04/ 9.2E-05/ 6.4E-01 |
| | G/G | 34 | 0.106 | 30 | 0.132 | 4 | 0.042 | 116 | 0.077 | 1.7 [1.1;2.6]/ 2.4 [1.5;3.9]/ 0.5 [0.2;1.5] | 1.1E-02/ 1.4E-04/ 2.6E-01 |
| | G allele | | 0.340 | | 0.377 | | 0.253 | | 0.268 | 1.4 [1.081;1.81]/ 1.7 [1.246;2.23]/ 0.9 [0.574;1.49] | 1.0E-02/ 1.0E-03/ 7.5E-01 |
| rs7089424 (*ARID5B*) | | | | | | | | | | | |
| | T/T | 126 | 0.391 | 82 | 0.361 | 44 | 0.463 | 683 | 0.453 | | |
| | G/T | 151 | 0.469 | 109 | 0.480 | 42 | 0.442 | 673 | 0.446 | 1.2 [0.9;1.6]/ 1.3 [1.0;1.8]/ 1.0 [0.6;1.5] | 1.4E-01/ 5.5E-02/ 8.9E-01 |
| | G/G | 45 | 0.140 | 36 | 0.159 | 9 | 0.095 | 152 | 0.101 | 1.6 [1.1;2.4]/ 2.0 [1.3;3.0]/ 0.9 [0.4;1.9] | 1.6E-02/ 2.0E-03/ 8.2E-01 |
| | G allele | | 0.374 | | 0.399 | | 0.316 | | 0.324 | 1.2 [0.964;1.59]/ 1.4 [1.047;1.86]/ 1.0 [0.616;1.50] | 9.4E-02/ 2.3E-02/ 8.6E-01 |
| rs3731217 (*CDKN2A*) | | | | | | | | | | | |
| | G/G | 2 | 0.006 | 1 | 0.004 | 1 | 0.011 | 32 | 0.021 | | |
| | G/T | 60 | 0.186 | 38 | 0.167 | 22 | 0.232 | 379 | 0.252 | 2.5 [0.6;10.8]/ 3.2 [0.4;24.1]/ 1.9 [0.2;14.2] | 2.1E-01/ 2.6E-01/ 7.8E-01 |
| | T/T | 260 | 0.807 | 188 | 0.828 | 72 | 0.758 | 1092 | 0.727 | 3.8 [0.9:16.0]/ 5.5 [0.7;40.6]/ 2.1 [0.3;15.7] | 6.8E-02/ 9.4E-02/ 6.9E-01 |
| | T allele | | 0.900 | | 0.912 | | 0.874 | | 0.853 | 1.6 [1.056;2.31]/ 1.8 [1.103;2.89]/ 1.2 [0.640;2.22] | 2.5E-02/ 1.7E-02/ 5.8E-01 |
| rs2239633 (*CEBPE*) | | | | | | | | | | | |
| | T/T | 57 | 0.177 | 33 | 0.145 | 24 | 0.253 | 307 | 0.203 | | |
| | C/T | 173 | 0.537 | 122 | 0.537 | 51 | 0.537 | 773 | 0.512 | 1.2 [0.9;1.7]/ 1.5 [1.0; 2.2]/ 0.8 [0.5;1.4] | 2.6E-01/ 6.4E-02/ 5.1E-01 |
| | C/C | 92 | 0.286 | 72 | 0.317 | 20 | 0.211 | 430 | 0.285 | 1.2 [0.8;1.6]/ 1.6 [1.0;2.4]/ 0.6 [0.3;1.1] | 4.4E-01/ 4.7E-02/ 1.0E-01 |
| | C allele | | 0.555 | | 0.586 | | 0.479 | | 0.541 | 1.0 [0.834;1.35]/ 1.2 [0.904;1.59]/ 0.8 [0.526;1.21] | 6.2E-01/ 2.1E-01/ 2.8E-01 |

have been made to set up an international consortium to also identify low-penetrance susceptibility alleles.[13] The identification and analysis of such low-penetrance variants may further facilitate the pathogenetic understanding of pediatric and adult ALL. Last, but not least, it has to be kept in mind that ethnic background may influence the risk and thus the hitherto reported risk factors may not be valid for all regions of the world.[14]

*Thomas Burmeister,[1]* Gesine Bartels,[1] Daniela Gröger,[1] Heiko Trautmann,[2] Stefan Schwartz,[1] Klaus Lenz,[3] Carola Tietze-Bürger,[4] Andreas Viardot,[5] Ralph Wäsch,[6] Heinz-August Horst,[2] Richard Reinhardt,[7] Nicola Gökbuget,[8]*

*Dieter Hoelzer,[8] Michael Kneba,[2] and Monika Brüggemann[2]*

*[1]Charité, Med. Klinik für Hämatologie, Onkologie und Tumorimmunologie, Berlin; [2]Christian-Albrechts-Universität zu Kiel, UKSH, Med. Klinik II, Kiel; [3]Charité, CBF, Institut für Biometrie und Klinische Epidemiologie, Berlin; [4]Charité, Stem Cell Facility, Berlin; [5]Universität Ulm, Klinik für Innere Medizin III, Ulm; [6]Albert-Ludwigs-Universität Freiburg, Klinik für Innere Medizin I, Freiburg; [7]Max-Planck-Genomzentrum, Köln; and [8]Goethe-Universität Frankfurt, Med. Klinik II, Frankfurt/Main, Germany*

*Correspondence: thomas.burmeister@charite.de*
*doi:10.3324/haematol.2013.090720*

*Acknowledgments: the authors thank all involved clinics and patients of the GMALL study group for their support and all involved laboratory technicians for their work. The authors thank Prof. Rajiv Kumar (DKFZ, Heidelberg, Germany) for communicating detailed rs3731217 genotype results of the control group. This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk.*

*Funding: TB was supported by grant 2011-05 from the Alfred & Angelika Gutermuth Foundation, Frankfurt and by grant 12/09f from the German José Carreras Leukemia Foundation. Funding for the project was provided by the Wellcome Trust under award 076113 and 085475.*

*Key words: acute lymphoblastic leukemia, adult-onset, risk factors, germline variants, IKFZ1, ARID5B, CEBPE.*

*Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.*

## References

1. Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B, Sheridan E, et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. Nat Genet. 2009;41(9):1006-10.
2. Treviño LR, Yang W, French D, Hunger SP, Carroll WL, Devidas M, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. Nat Genet. 2009;41(9):1001-5.
3. Sherborne AL, Hosking FJ, Prasad RB, Kumar R, Koehler R, Vijayakrishnan J, et al. Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. Nat Genet. 2010;42(6):492-4.
4. Prasad RB, Hosking FJ, Vijayakrishnan J, Papaemmanuil E, Koehler R, Greaves M, et al. Verification of the susceptibility loci on 7p12.2, 10q21.2, and 14q11.2 in precursor B-cell acute lymphoblastic leukemia of childhood. Blood. 2010;115(9):1765-7.
5. Burmeister T, Meyer C, Schwartz S, Hofmann J, Molkentin M, Kowarz E, et al. The MLL recombinome of adult CD10-negative B-cell precursor acute lymphoblastic leukemia: results from the GMALL study group. Blood. 2009;113(17):4011-5.
6. Burmeister T, Gökbuget N, Schwartz S, Fischer L, Hubert D, Sindram A, et al. Clinical features and prognostic implications of TCF3-PBX1 and ETV6-RUNX1 in adult acute lymphoblastic leukemia. Haematologica. 2010;95(2):241-6.
7. Brüggemann M, Raff T, Flohr T, Gökbuget N, Nakao M, Droese J, et al. Clinical significance of minimal residual disease quantification in adult patients with standard-risk acute lymphoblastic leukemia. Blood. 2006;107(3):1116-23.
8. Burmeister T, Marschalek R, Schneider B, Meyer C, Gökbuget N, Schwartz S, et al. Monitoring minimal residual disease by quantification of genomic chromosomal breakpoint sequences in acute leukemias with MLL aberrations. Leukemia. 2006;20(3):451-7.
9. Gast A, Bermejo JL, Flohr T, Stanulla M, Burwinkel B, Schrappe M, et al. Folate metabolic gene polymorphisms and childhood acute lymphoblastic leukemia: a case-control study. Leukemia. 2007;21(2):320-5.
10. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78.
11. van Vlierberghe P, Palomero T, Khiabanian H, van der Meulen J, Castillo M, van Roy N, et al. PHF6 mutations in T-cell acute lymphoblastic leukemia. Nat Genet. 2010;42(4):338-42.
12. Peyrouze P, Guihard S, Grardel N, Berthon C, Pottier N, Pigneux A, et al. Genetic polymorphisms in ARID5B, CEBPE, IKZF1 and CDKN2A in relation with risk of acute lymphoblastic leukaemia in adults: a Group for Research on Adult Acute Lymphoblastic Leukaemia (GRAALL) study.[letter]. Br J Haematol 2012;159(5):599-613.
13. Sherborne AL, Hemminki K, Kumar R, Bartram CR, Stanulla M, Schrappe M, et al. Rationale for an international consortium to study inherited genetic susceptibility to childhood acute lymphoblastic leukemia. Haematologica. 2011;96(7):1049-54.
14. Xu H, Cheng C, Devidas M, Pei D, Fan Y, Yang W, 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(7):751-7.