Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia

E Ellinghaus1,19, M Stanulla2,19, G Richter1, D Ellinghaus1, G te Kronnie3, G Cario2, G Cazzaniga4, M Horstmann5, R Panzer Grümayer6, H Cavé7, J Trka8, O Cinek9, A Teigler-Schlegel10, A ELSharawy1, R Häslter1, A Nebel1, B Meissner2, T Bartram2, F Lescai11, C Franceschi12, M Giordan13, P Nürnberg13, B Heinzow14,15, M Zimmermann16, S Schreiber1,17,18,19, M Schrappe2,19 and A Franke1,19

Acute lymphoblastic leukemia (ALL) is a malignant disease of the white blood cells. The etiology of ALL is believed to be multifactorial and likely to involve an interplay of environmental and genetic variables. We performed a genome-wide association study of 355,750 single-nucleotide polymorphisms (SNPs) in 474 controls and 419 childhood ALL cases characterized by a t(12;21)(p13;q22) – the most common chromosomal translocation observed in childhood ALL – which leads to an ETV6-RUNX1 gene fusion. The eight most strongly associated SNPs were followed-up in 951 ETV6-RUNX1-positive cases and 3061 controls from Germany/Austria and Italy, respectively. We identified a novel, genome-wide significant risk locus at 3q28 (TP63, rs17505102, P_CMS = 8.94 × 10⁻⁹, OR = 0.65). The separate analysis of the combined German/Austrian sample only, revealed additional genome-wide significant associations at 11q11 (OR8B8, rs1945213, P = 9.14 × 10⁻¹¹, OR = 0.69) and 8p21.3 (near INTS10, rs929590, P = 6.12 × 10⁻⁹, OR = 1.36). These associations and another association at 11p11.2 (PTPRJ, rs3942852, P = 4.95 × 10⁻⁷, OR = 0.72) remained significant in the German/Austrian replication panel after correction for multiple testing. Our findings demonstrate that germline genetic variation can specifically contribute to the risk of ETV6-RUNX1-positive childhood ALL. The identification of TP63 and PTPRJ as susceptibility genes emphasize the role of the TP53 gene family and the importance of proteins regulating cellular processes in connection with tumorigenesis.

Leukemia (2012) 26, 902–909; doi:10.1038/leu.2011.302; published online 11 November 2011

Keywords: genome-wide association study; childhood acute lymphoblastic leukemia; TP63

INTRODUCTION

Acute lymphoblastic leukemia (ALL) is characterized by the malignant clonal proliferation of lymphoid cells that are blocked at an early stage of differentiation. More than 60% of patients diagnosed with ALL are children below the age of 15 years.1,2,8,9 The annual incidence rates of childhood ALL vastly worldwide between one and four new cases per 100,000 children younger than 15 years, with a peak incidence at about 2-5 years of age.3,5

As for most other human malignancies not underlying a clear hereditary genetic trait, the etiology of childhood ALL is believed to be multifactorial, including environmental risk factors like radiation and other toxins as well as genetic variables.6,7 Different subtypes of childhood ALL are classifiable according to either immunophenotypic (for example, precursor B-cell vs T-cell lineage ALL) or a constantly growing number of recurrent somatic genetic aberrations that have been characterized during the last decades.1,2,8,9 The two major genetic subtypes of childhood ALL are characterized by either hyperdiploidy (>50 chromosomes per leukemic cell) or the chromosomal translocation t(12;21)(p13;q22) leading to an ETV6–RUNX1 (TEL–AML1) gene fusion.10 Hyperdiploid and ETV6–RUNX1-rearranged childhood ALL account for ~25% and 22% of the entire childhood ALL population, respectively. A minority of childhood ALL cases (<5%) is related to specific genetic conditions such as Down’s syndrome or disorders associated with impaired DNA repair capacities (for example, ataxia teleangiectasia, Nijmegen breakage syndrome or Bloom’s
syndrome). Recently, two genome-wide association studies (GWAS) on childhood ALL performed in the UK and the US have identified 13 susceptibility loci.1,12 The UK study screened around 291 000 single-nucleotide polymorphisms (SNPs) in 907 ALL cases (a mixture of different subtypes) in a meta-analysis without independent replication while the US study investigated nearly 308 000 SNPs in 317 ALL cases (also a mixture) after quality control. The two highest scoring loci were detected by both studies and subsequently verified by others:13 IKZF1, a transcription regulator of lymphoid cell differentiation, and ARID5B, a member of the AT-rich interaction domain family of transcription factors.

We report here the results of a genome-wide association study conducted with patients displaying a subtype of childhood ALL characterized by the chromosomal translocation which leads to an ETV6–RUNX1 gene fusion. Thus, to the best of our knowledge, our study is the first genome-wide approach to susceptibility loci primarily addressing one of several well-defined and specific subgroups of the heterogenous disease entity ‘childhood ALL’.

PATIENTS AND METHODS

Recruitment of patients and healthy controls

Samples were organized in panels that corresponded to the successive steps of the present study. All individual panels (A–D, Supplementary Table S1) were independent from each other.

All patients included in panel A, 206 patients from panel B, and all patients from panels C and D were enrolled into the Austrian-German-Italian-Swiss multicenter clinical trial AIEOP–BFM ALL 2000 on treatment of childhood ALL after informed consent had been obtained. Patients were diagnosed and treated at one of 121 participating study centers in Austria, Germany and Italy between 1999 and 2008. 200 ALL patients from panel B were enrolled into two subsequent clinical trials of the German COALL study group, COALL-06-97 and COALL-07-03 between 1997 and 2008 at 19 German treatment centers. To generate the extended panel B with 664 patients, the 406 patients from panel B were complemented with an additional 258 patients enrolled in the years 2009 and 2010 into the Austrian-German-Italian-Swiss multicenter clinical trials AIEOP–BFM ALL 2000 or its predecessor, the ongoing trial AIEOP–BFM ALL 2009 after informed consent had been obtained. Diagnosis was based on cytomorphology (FAB criteria) and cytochemistry when >25% of lymphoblastic cells were present in the bone marrow, or when lymphoblasts were present in the peripheral blood or cerebrospinal fluid. Flow cytometric immunophenotyping was performed according to consensus protocols based on the guidelines proposed by the European Group for the immunological characterization of Leukemias. The presence or absence of ETV6–RUNX1, BCR–ABL and MLL–AF4 fusion gene transcripts was analyzed by a multiplex PCR assay enabling the detection of M-BCR–ABL, m-BCR–ABL, ET6-V6-RUNXI and MLL–AF4 fusion transcripts in a single PCR reaction. Positive results were confirmed by interphase fluorescence in situ hybridization in the majority of the patients. Written, informed consent was obtained from all study participants or their legal guardians to use sparse diagnostic specimens for research purposes and all protocols were approved by the respective institutional ethical review committees.

Bone marrow samples were obtained at initial diagnosis and at consecutive follow-up time points during therapy, mainly after induction and consolidation. Mononuclear cells were isolated by Ficoll-Paque gradient centrifugation (Pharmacia, Freiburg, Germany) from bone marrow samples followed by extraction of high-molecular weight DNA according to standardized protocols. Quality and quantity of genomic DNA was determined by spectrophotometry. In the present study only remission DNA samples followed by extraction of high-molecular weight DNA according to standardized protocols. Quality and quantity of genomic DNA was determined by spectrophotometry. In the present study only remission DNA was ligated to their respective Nsp I adaptor and then labeled with biotin. Labeled DNA was combined with hybridization mix and then injected into array. Arrays were hybridized for 18-22 h at 50 °C. DNA samples were recovered from arrays and washed and stained by using Affymetrix FS450 fluidic stations. Stained arrays were scanned using Affymetrix GeneChip Scanner 3000 7G.

Raw image files were converted into cel-files by Affymetrix genotyping console. A preliminary QC call rate of 86% was used to pass arrays for further data analysis. Passed arrays were clustered in the same batches which were processed together in the lab. Genotypes were assigned using the BRLMM-p algorithm. There were 12 positive control samples run in each project. The positive controls were from a CEPH trios family: NA12740 – daughter (six repeats), NA12750 – father (four repeats), and NA12751 – mother (four repeats). This set of controls allowed to calculate experimental reproducibility, trio accuracy and HapMap concordance. The average call rate was above 96%, and the average BRLMM-p call rate, average reproducibility (controls), average HapMap concordance (controls) and average trio accuracy were above 99%.

Samples with more than 5% missing genotypes, who showed excess genetic dissimilarity to the other subjects or who showed evidence for cryptic relatedness to other study participants (Supplementary Figure S1) were removed. These quality control measures left 419 ETV6–RUNX1-positive ALL patients and 474 healthy control samples for inclusion in screening panel A. All gender assignments could be verified by reference to the proportion of heterozygous SNPs on the X chromosome. Before analysis, all SNPs that had a low genotyping rate (<95% in cases or controls), were monomorphic or rare (minor allele frequency < 2% in cases or controls), or deviated from Hardy-Weinberg equilibrium in the control population (P_HWE < 0.001) were excluded (n = 88 066; 19.8% of all SNPs).

SNP genotyping for genome-wide screening

The genotyping for the GWAS samples (panel A), which was part of the German NGFN GWAS initiative funded by the NGFN, was performed as a service project by Affymetrix (South San Francisco, CA, USA) using the genome-wide human SNP array 5.0 (500 k) with 443 816 markers. The array is based on an assay termed whole-genome sampling analysis developed for highly multiplexed SNP genotyping of complex DNA. This method reproducibly amplifies a subset of the human genome through a single primer amplification reaction using restriction enzyme digested and adapter-ligated human genomic DNA.

In brief, 5 μl of genomic DNA samples at 50 ng/μl were aliquoted to the corresponding wells of two 96-well plates. The first run of samples was processed as an entire plate. In the lab, transfers were made with a 12-channel pipette, reducing the risk of sample tracking errors. One plate was digested with Nsp I and the other plate was digested with Sty I. The reaction was incubated at 37 °C for 2 h and at 65 °C for 20 min to deactivate the enzyme. The digested DNA was then ligated to their respective Nsp I adaptor and Sty I adaptor. The ligated product was then PCR-amplified using a common primer. Both Nsp I PCR product and Sty I PCR product were combined, and then purified by ethanol precipitation in combination with membrane filter plate. Purified PCR product was further fragmented with DNase I and then labeled with biotin. Labeled DNA was combined with hybridization mix and then injected into array. Arrays were hybridized for 18-22 h at 50 °C. DNA samples were recovered from arrays and washed and stained by using Affymetrix FS450 fluidic stations. Stained arrays were scanned using Affymetrix GeneChip Scanner 3000 7G.
**RESULTS**

Associations with ETV6–RUNX1 subtype of childhood ALL

Novel significant disease associations that withstood Bonferroni correction (adjusted significance threshold \(= 5.75 \times 10^{-6}\) (0.05/871)) in the combined analysis of replication panels B and C were obtained for rs17505102, an intronic SNP in the gene encoding the tumor protein p63 isoform 3 (TP63; also known as tumor protein p73-like: TP73L) on chromosome 3q28 (\(P_{\text{CMH}} = 4.87 \times 10^{-7}\), OR = 0.63, 95% CI = 0.52-0.75) and for the intronic SNP rs3942852, located in the gene PTPRJ (protein tyrosine phosphatase, receptor type, J) on chromosome 11p11.2 (\(P_{\text{CMH}} = 2.54 \times 10^{-4}\), OR = 0.77, 95% CI = 0.68-0.89) (Table 1 and Figure 1). Although the associations of these loci with ETV6–RUNX1-positive ALL did not achieve genome-wide significance (conventional level of \(P < 5.0 \times 10^{-8}\)) in the discovery panel, we obtained genome-wide significance for rs17505102 (\(P_{\text{CMH}} = 8.94 \times 10^{-8}\)) within the TP63 gene in the overall sample of panels A through C comprising 1370 ETV6–RUNX1-positive ALL cases and 3535 healthy controls. For rs3942852, the combined analysis of the overall sample yielded \(P_{\text{CMH}} = 1.00 \times 10^{-6}\).

As the initial screening (discovery panel A) was performed with German samples only, we additionally analyzed the follow-up SNPs in the German/Austrian replication panel B exclusively to identify potential loci that are more specific for northern Europeans as compared with those that confer susceptibility in the European population in general. Apart from the two novel shared susceptibility loci (rs17505102 and rs3942852), we found another two associated SNPs that remained significant after correction for multiple testing: rs1945213 on chromosome 11q11 within the OR8U8 gene (olfactory receptor, family 8, subfamily U, member 8; \(P = 1.63 \times 10^{-7}\)) and rs920590 on 8p21.3, 24 kb upstream of the Integrator complex subunit 10 (\(P = 1.05 \times 10^{-7}\)). Furthermore, three of these associations (rs17505102, rs1945213 and rs920590) were genome-wide significant in the combined analysis of the German/Austrian panels A and B comprising 1083 cases and 2956 controls. For rs3942852 we obtained a \(P\)-value of \(4.95 \times 10^{-7}\) in the combined analysis (Table 1 and Figure 1). Of the newly identified susceptibility loci especially the PTPRJ gene and the tumor protein 63 are very interesting candidates, taking account of their functional role and strong homology with the tumor suppressor gene TP53, respectively. As the marker rs17505102 is located in the first intron of the TP63 gene, we sequenced exons 1-3 in 47 ALL cases to determine whether the lead SNP might tag one or more potential causative coding polymorphisms via linkage disequilibrium. However, we did not identify any coding SNP in these regions. Additionally, we checked NCBi’s dbSNP\(^{19}\) build 134 for reported and validated, non-synonymous coding SNPs (\(\geq 1\%\) frequency) within the TP63 gene, but found no SNP entries satisfying these criteria.

Using a publicly available resource (web-based tool SNPexp v1.1\(^{17}\)), we were not able to identify any significant correlation - so-called eQTL effect – between the genotypes at the lead SNPs rs17505102, rs1945213, rs3942852 and rs920590 and the expression levels of the nearby located genes (TP63, PTPRJ and INTS10) in EBV-transformed lymphoblastoid cell lines from the HapMap samples (available from the GENEVAR project\(^{19}\)). For OR8U8 no expression data were available. However, it is noteworthy that the SNP rs3942852, located in the first intron of the PTPRJ gene,
| Chr position (bp) | dSNP ID | Nearby gene | Panel A - GWAS | Panel B extended | Panel C | Panel B extended + C | Panel C + Bext | Combined analysis |
|------------------|---------|-------------|----------------|-----------------|---------|---------------------|----------------|------------------|
| 3                | rs1750102 | TP63        | A1             | A1              | 2.09e-05 | 4.87x 10^-3       | 0.44           | 8.94e 10^-3      |
| 11               | rs1945213 | GRB18       | 0.105          | 10              | 2.79e-05 | 3.30              | 2.41           | 2.21x 10^-4      |
| 5                | rs932247  | ETV6        | 0.105          | 10              | 2.41e-05 | 3.29              | 2.21           | 2.21x 10^-4      |
| 11               | rs2948252 | RUNX1       | 0.01           | 10              | 2.54e-08 | 5.60x 10^-12      | 0.21           | 1.35x 10^-6      |
| 11               | 48071665  | RUNX1       | 0.01           | 10              | 2.21e-05 | 0.54              | 2.21x 10^-6    | 0.54x 10^-6      |
| 14               | rs1756960 | CCM1        | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 7                | rs207954  | LCP3        | 0.01           | 10              | 2.41e-08 | 5.60x 10^-12      | 0.21           | 1.35x 10^-6      |
| 15               |         |             |                |                 |         |                    |                |                  |
| 10               | rs2007930 | PTEN        | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 8                | rs1920590 | PTEN        | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 4                | rs282708  | -           | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 9                | rs196483  | CDCA1       | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 10               | rs2192205 | CDCA1       | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 10               | rs1857366 | AFN1       | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 10               | rs7734914 | -           | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 10               | rs105535  | -           | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 6                | rs7378636 | -           | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 11               | rs334604  | KIAA1      | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 11               | rs1100293 | KIAA1      | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 8                | rs630662  | RSP2       | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 10               | rs9041474 | RSP2       | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 13               | rs7336133 | LAMA8      | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 8                | rs845574  | -           | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 5                | rs1742910 | PDE4B      | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 6                | rs646197  | -           | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 6                | rs6501152 | AGT        | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 14               | rs6837005 | ETV6       | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 5                | rs6445754 | ETV6       | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 5                | rs657295  | ETV6       | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 14               | rs901921  | DNTM1      | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 5                | rs908882  | CCDC130    | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |
| 7                | rs901780  | CCDC130    | 0.01           | 10              | 2.12e-05 | 0.25              | 2.21           | 0.54x 10^-6      |

Abbreviations: AF, allele frequency; Chr, chromosome; CI, confidence interval; GWAS, genome-wide association study; OR, odds ratio; SNP, single-nucleotide polymorphism. *The 12 SNPs that were not genotyped in the combined replication analysis (panels B and C) are highlighted in bold. **SNP associations that remained significant after correction for multiple testing (combined analysis) are highlighted in bold. SNP associations that remained significant after correction for multiple testing (combined analysis) are highlighted in bold. **SNP associations that remained significant after correction for multiple testing (combined analysis) are highlighted in bold.

Identification of risk loci in ALL subtype Leukemia (2012) 902 - 909
resides within a transcription factor binding site of BAF155 (SMARCC1), BAF170 (SMARCC2) and the proto-oncogene JunD assayed by CHiPseq in HeLa-S3 cells (UCSC Browser hg18). Hence, further experiments and analyses are required to elucidate the mechanism by which the SNPs rs17505102, rs1945213, rs3942852 and rs920590 affect the risk of ALL.

Associations with childhood ALL independent of \( \text{ETV6} - \text{RUNX1} \) subtype

In order to assess whether our findings were specifically associated with the \( \text{ETV6} - \text{RUNX1} \)-rearranged subtype, we genotyped the 20 top SNPs displayed in Table 1 in a second German replication cohort (panel D) consisting of 326 \( \text{ETV6} - \text{RUNX1} \)-negative childhood ALL cases. We analyzed these cases together with the same 1682 controls of panel B and obtained genotypes of 19 SNPs after quality control. Of the 19 SNPs only rs7738636 on chromosome 6q14.1 was significantly associated (\( P = 2.55 \times 10^{-3} \)). The combined analysis of the replication panels B/Bextended through D yielded only a marginally stronger association (\( P_{\text{CMH}} = 1.42 \times 10^{-3} \)) for rs7738636 in the combined sample (panels A-D) comprising 1438 cases and 2735 controls. Suggestive evidence of association (\( P < 0.05 \)) was found for rs6428370, 1q31.3 (rs6428370), KCNMB2, PARD3 and C12orf5 (Table 2).

**DISCUSSION**

In this study, we were able to identify attractive, novel risk loci for ALL, even with a small screening sample - compared with other GWAS studies and diseases - of only 419 patients. This success might be due to the exclusive selection of patients with the homogenous phenotype of \( \text{ETV6} - \text{RUNX1} \)-rearranged childhood ALL. Given their interesting biological function, we below discuss the relevance of \( \text{TP63} \) and \( \text{PTPRJ} \) in ALL disease etiology.

The \( \text{TP63} \) gene - a member of the \( \text{TP53} \) gene family - codes for p63 which has essential roles in embryonic development. \( \text{TP63} \) contains two transcriptional start sites leading to p63 isoforms either containing (TAp63) or lacking (\( \text{W} \)Np63) the trans-activation domain.\(^{21}\) TP63-null mice die cancer-free within a few hours of birth, showing severe loss of limbs and a wide range of epithelial structures including skin, prostate, breast and urothelia.\(^{22,23}\) This suggests that p63 acts as an important regulator of stemness in developing epithelia.\(^{24,25}\) In humans, TP63 mutations are associated with ectodermal dysplasia, cleft lip or palate and limb malformations, but not with a higher tumor incidence.\(^{21,26}\) In line with these observations, the TP63 coding region is found only respectively, in panels A through D in order to confirm the previously reported loci and as a positive control for our experiment. We observed genome-wide significant association (\( P < 5 \times 10^{-8} \)) for IKZF1, DDC, ARID5B and CEBPE in the combined sample (panels A-D) comprising 1438 cases and 2735 controls. Suggestive evidence of association (\( P < 0.05 \)) was found for \( \text{SIAT7C} \), 1q31.3 (rs6428370), KCNMB2, PARD3 and C12orf5 (Table 2).

**Figure 1.** Regional plots of the confirmed associations at \( \text{TP63} \), \( \text{OR8U8} \), on 8p21.3 and at \( \text{PTPRJ} \). The regional plots of the negative decadic logarithm of the \( P \)-values obtained in the GWAS (panel A) are shown. Panel A was imputed with CEU haplotypes generated by the 1000 Genomes Project (August 2010 release) as reference. For the central lead SNP of each plot, the combined \( P \)-value of panels A and B (rs1945213, rs920590 and rs3942852) or panels A through C (rs17505102) is indicated. The magnitude of linkage disequilibrium (LD) with the central SNP measured by \( r^2 \) is reflected by the color of each SNP symbol (for color coding, see upper right corner of each plot). Recombination activity (in centimorgans (cM) per Mb) is depicted by a blue line. Positions are given as NCBI’s build 36 coordinates. For details, see Table 1.

Associations of previously reported susceptibility-conferring SNPs

Additionally, we genotyped the susceptibility-conferring SNPs identified by Papaemmanuil et al.\(^{11}\) and Trevino et al.,\(^{12}\) respectively, in panels A through D in order to confirm the previously reported loci and as a positive control for our experiment. We observed genome-wide significant association (\( P < 5 \times 10^{-8} \)) for IKZF1, DDC, ARID5B and CEBPE in the combined sample (panels A-D) comprising 1438 cases and 2735 controls. Suggestive evidence of association (\( P < 0.05 \)) was found for \( \text{SIAT7C} \), 1q31.3 (rs6428370), KCNMB2, PARD3 and C12orf5 (Table 2).
| Chr position (bp) | dbsNP ID | Gene name | A1 | A2 | Panel A Germany | Panel B Germany | Panel C Italy | Panel A + B + C Combined analysis | Panel D Germany | Panel A + B + C + D Combined analysis |
|------------------|----------|-----------|----|----|----------------|----------------|-------------|---------------------------------|----------------|-----------------------------------|
|                  |          |           |    |    | ETV6 - RUNX1 pos. 419 cases | ETV6 - RUNX1 pos. 419 cases | ETV6 - RUNX1 pos. 287 cases | ETV6 - RUNX1 pos. 287 cases | ETV6 - RUNX1 pos. 287 cases | ETV6 - RUNX1 pos. 287 cases |
|                  |          |           |    |    | 474 controls | 406 cases | 579 controls | 1112 cases | 326 cases | 2735 controls | 1438 cases |
|                  |          |           |    |    | 0.94 | 0.94 | 0.87 | 0.93 | 0.83 | 0.87 | 0.87 |
|                  |          |           |    |    | 0.04 | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |

Identification of risk loci in ALL subtype
Identification of risk loci in ALL subtype

E Ellinghaus et al

rarely mutated in human cancers.\textsuperscript{27,28} To date, recurrent somatic TP63 mutations have only been described for non-small cell lung cancer as well as chronic myeloid leukemia\textsuperscript{29,30} and TP63 germline genetic variation (rs710521) has only been associated with urinary bladder cancer, so far.\textsuperscript{31} However, mice heterozygous for TP63 mutations develop malignant lesions, demonstrating that TP63 can act as a tumor suppressor and mice heterozygous for mutations in both TP53 and TP63 display higher tumor burden and metastasis compared with mice only heterozygous for TP53 mutations.\textsuperscript{32} Of particular interest to our study, TAp63 was described to be involved in an antiapoptotic pathway regulating normal B and chronic lymphocytic leukemia cell survival in a CD74-dependent manner.\textsuperscript{33} In chronic lymphocytic leukemia, TAp63 expression upregulates VLA-4 integrin expression leading to augmented migration and homing of chronic lymphocytic leukemia cells to the bone marrow.\textsuperscript{34} Also, it was found that thymomas and certain B-cell lymphomas express high levels of TP63, suggesting a role of TP63 in these cancers.\textsuperscript{35} TP63 is a receptor type protein tyrosine phosphatase involved in the regulation of cellular processes including cell growth, differentiation, mitotic cycle and oncogenic transformation. Somatic mutations of TP63 have been described in a wide spectrum of cancers like thyroid carcinomas, colon cancer, breast and lung cancer.\textsuperscript{35-38} Furthermore, a specific TP63 haplotype was shown to be associated with the risk of breast cancer demonstrating that common TP63 variants may act as breast cancer susceptibility alleles.\textsuperscript{39} Interestingly, the TP63 gene is expressed at high levels in hematopoietic cells\textsuperscript{35-37,40} Of particular interest to the results reported in our study, mice with a constitutively deleted transmembrane domain allele of TP63 display a partial peripheral B-cell developmental block.\textsuperscript{41} Heterozygous mice do not show any notable phenotype.

Although the current knowledge on TP63 and TP6R may provide some evidence for their role in tumorigenesis as well as B-cell development and/or maintenance, more precise mechanisms how germline genetic variation at the TP63 and TP6R loci may modulate susceptibility to precursor B-cell leukemia remain speculative and underscore the need for functional studies.

Apart from our novel findings, our study could clearly confirm the association of childhood ALL with the four previously published susceptibility loci IQF1, DDC, ARID5B and CEBPE. Nominal significant P-values (P < 0.05) were obtained for SIAL7C, 1q31.3, KCNMB2, PARD3 and C12orf5. Although our study panel had more than 80% power to detect all of the SNPs/loci published by Papaemmanuil et al.\textsuperscript{11,12} respectively, we could not replicate the loci RYR2, OR2C3, KCN4, 6q24.1, KRTHB5, 18p11.32 and ZNF230, which therefore require further investigations.

Overall, our findings demonstrate that germline genetic variation can specifically contribute to the risk of specific ALL subtypes, in this case ETV6-RUNX1-positive childhood ALL. Our results also suggest that germline genetic variation can act as a risk factor for childhood ALL in general or functions restricted to specific subtypes, only.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The authors wish to thank all ALL patients, families and physicians for their cooperation. Also, we wish to thank T Wesse, T Henke, M Wittig, R Vogler, L Tittmann and R Fedders for expert technical help. We are grateful to C Becker, R Nemiec and C Kluck for genotyping the samples of GWAS panel A, Prof Dr T Wienker and Dr M Steffens (IMBIE, University of Bonn) are acknowledged for performing the quality control of the GWAS control data set. This study was supported by the German Ministry of Education and Research (BMBF) through the National Genome Research Network (NGFN), the popgen biobank, the Federal Radiation Protection Agency (project no. 360930013), the Deutsche Krebshilfe and the Madeleine

Table 2 (Continued)
Identification of risk loci in ALL subtype

E Ellinghaus et al

909

Schickedanz Kinderkrebs-Stiftung. The project received infrastructure support through the DFG excellence cluster Inflammation at Interfaces. This study was conducted within the frame of the International BFM Study Group.

AUTHOR CONTRIBUTIONS

EE, AF and GR performed the SNP selection, genotyping, data analysis and resequencing. EE prepared figures and tables. DE prepared regional plots and helped with data analysis and quality control. AE, HH, AN and MW helped with data analysis. PN coordinated the Affymetrix chip genotyping. MS, G Car and M Sch coordinated the recruitment and collected the phenotype data. GK, G Caz, JT, OC, HC, RPG, BM, TB, CF and MH recruited the additional case-control panels. AT-S, GK, MG, G Caz, and RPG were involved in phenotypic characterization of cases. AF, BH, G Car, M Sch, M St and SS designed and supervised the experiment. EE, M St and AF wrote the manuscript. All authors approved the final manuscript.

URLs

SNPexpl25 v1.1: http://app3.titan.uio.no/biotools/tool.php?app = snpexpl R script for Regional Association Plots: http://www.broadinstitute.org/science/projects/diabetes-genetics-initiative/ploting-genome-wide-association-results 1000 Genomes data: http://www.sph.umich.edu/csg/abecasis/MACH/download/

REFERENCES

1 Pui CH, Evans WE. Treatment of acute lymphoblastic leukemia. N Engl J Med 2006; 354: 166 - 178.
2 Stanulla M, Schrappe M. Treatment of childhood acute lymphoblastic leukemia. Semin Hematol 2009; 46: 52 - 63.
3 Redaelli A, Laskin BL, Stephens JM, Botteman MF, Pashos CL. A systematic literature review of the clinical and epidemiological burden of acute lymphoblastic leukaemia (ALL). Eur J Cancer Care (Engl) 2005; 14: 53 - 62.
4 Linabery AM, Ross JA. Trends in childhood cancer incidence in the US (1992 - 2004). Cancer 2008; 112: 416 - 432.
5 Howard SC, Metzger ML, Willimas JA, Quintana Y, Pui CH, Robison LL et al. Childhood cancer epidemiology in low-income countries. Cancer 2008; 112: 461 - 472.
6 Belsen M, Kingsley B, Holmes A. Risk factors for acute leukemia in children: a review. Environ Health Perspect 2007; 115: 138 - 145.
7 Eden T. Aetiology of childhood leukaemia. Cancer Treat Rev 2010; 36: 286 - 297.
8 Teitell MA, Pandolfi PP. Molecular genetics of acute lymphoblastic leukemia. Annu Rev Pathol 2009; 4: 175 - 198.
9 Mullighan CG, Downing JR. Genome-wide profiling of genetic alterations in acute lymphoblastic leukemia: recent insights and future directions. Leukemia 2009; 23: 1209 - 1218.
10 Greaves M, Darwin and evolutionary tales in leukemia. The Ham-Wasserlcm Lecture. Hematol Am Soc Hematol Educ Program 2009; 2009: 3 - 12. PMID: 20088176.
11 Papaemmanoul 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: 1006 - 1010.
12 Trevino 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: 1001 - 1005.
13 Prasad RB, Hosking FJ, Vijayakrishnan J, Papaemmanoul 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: 1765 - 1767.
14 Krawczak M, Nikolaus S, von Eberstein H, Croucher FJ, El Mokhtari NE, Schreiber S. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. Community Genet 2006; 9: 55 - 61.
15 Dupont WD, Plummer WD. PS power and sample size program available for free on the Internet. Controlled Clin Trials 1997; 18: 274.
16 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559 - 575.
17 Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 2009; 84: 210 - 223.
18 Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C et al. Population genomics of human gene expression. Nat Genet 2007; 39: 1217 - 1224.

Supplementary Information accompanies the paper on the website of the Leukemia project (http://www.nature.com/leu)