Open

ORIGINAL ARTICLE

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

E Ellinghaus[1,19], M Stanulla[2,19], G Richter[1], D Ellinghaus[1], G te Kronnie[3], G Cario[2], G Cazzaniga[4], M Horstmann[5], R Panzer Grümayer[6], H Cavé[7], J Trka[8], O Cinek[9], A Teigler-Schlegel[10], A ElSharawy[1], R Häsler[1], A Nebel[1], B Meissner[2], T Bartram[2], F Lescai[11], C Franceschi[12], M Giordan[3], P Nürnberg[13], B Heinzow[14,15], M Zimmermann[16], S Schreiber[1,17,18,19], M Schrappe[2,19] and A Franke[1,19]

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, $P_{CMH} = 8.94 \times 10^{-9}$, OR = 0.65). The separate analysis of the combined German/Austrian sample only, revealed additional genome-wide significant associations at 11q11 (*OR8U8*, rs1945213, $P = 9.14 \times 10^{-11}$, OR = 0.69) and 8p21.3 (near *INTS10*, rs920590, $P = 6.12 \times 10^{-9}$, OR = 1.36). These associations and another association at 11p11.2 (*PTPRJ*, rs3942852, $P = 4.95 \times 10^{-7}$, 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.[1–3] The annual incidence rates of childhood ALL vary 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

[1]Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel, Germany; [2]Department of Pediatrics, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany, on behalf of the German Berlin-Frankfurt-Münster Study Group for Treatment of Childhood Acute Lymphoblastic Leukemia; [3]Department of Pediatrics, Laboratory of Pediatric Hematology/Oncology, University of Padua, Padua, Italy; [4]M. Tettamanti Research Center, Children's Hospital, University of Milan-Bicocca, Monza, Italy; [5]Clinic of Pediatric Hematology and Oncology, University Medical Center, and Research Center Children's Cancer Center, Hamburg, Germany; [6]St Anna Children's Hospital and Children's Cancer Research Institute, Vienna, Austria; [7]Department of Genetics, Hôpital Robert Debré, Paris, France; [8]Department of Pediatric Hematology/Oncology, Second Faculty of Medicine, Charles University Prague, Prague, Czech Republic; [9]Department of Pediatrics, Second Faculty of Medicine, Charles University Prague, Prague, Czech Republic; [10]Oncogenetic Laboratory, Department of Pediatric Hematology and Oncology, Justus-Liebig-University, Giessen, Germany; [11]Division of Research Strategy, University College London, London, UK; [12]Department of Experimental Pathology, University of Bologna, Bologna, Italy; [13]Cologne Center for Genomics, University of Cologne, Cologne, Germany; [14]State Social Services Agency Schleswig-Holstein, Kiel, Germany; [15]University of Notre Dame, Sydney Medical School, Sydney, Australia; [16]Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany; [17]Department of General Internal Medicine, University Hospital Schleswig-Holstein, Christian-Albrechts-University Kiel, Kiel, Germany and [18]Popgen Biobank, University Hospital Schleswig-Holstein, Christian-Albrechts-University Kiel, Kiel, Germany. Correspondence: Professor Dr rer. nat. A Franke, Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Schittenhelmstr. 12, D-24105 Kiel, Germany.
E-mail: a.franke@mucosa.de
or Professor Dr med. M Stanulla, Department of Pediatrics, University Hospital Schleswig-Holstein, Campus Kiel, Arnold-Heller-Str. 3, D-24105 Kiel, Germany.
E-mail: martin.stanulla@uk-sh.de
[19]These authors contributed equally to this work.
Received 9 June 2011; revised 3 September 2011; accepted 21 September 2011; published online 11 November 2011

syndrome).[6,7] Recently, two genome-wide association studies (GWAS) on childhood ALL performed in the UK and the US have identified 13 susceptibility loci.[11,12] 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*, *ETV6–RUNX1* 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 spare 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 obtained at follow-up time points after induction or consolidation was analyzed.

The German healthy control individuals of panel A and B were obtained from the popgen biobank.[14] The healthy Italian controls of panel C were recruited through the blood bank of the San Gerardo Hospital at Monza, Italy. Written and informed consent was obtained from all study participants and all protocols were approved by the institutional ethical review committees of the participating centers. To account for their heterogeneous sources, all DNA samples were first quality-controlled by gel electrophoresis.

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.

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 trio 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 sample ($P_{HWE} \leqslant 0.001$) were excluded ($n = 88 066$; 19.8% of all SNPs).

SNPlex, TaqMan and Sequenom genotyping
All downstream genotyping (panels B through D) was performed with SNPlex and TaqMan technologies (Applied Biosystems, Foster City, CA, USA) or the Sequenom platform using an automated laboratory setup and all process data were written to and administered by a database-driven laboratory information management system.

Of 100 selected SNPs 87 SNPs passed quality control in both replication samples. These SNPs had a high call rate (>95% in cases or controls), were not monomorphic or rare (minor allele frequency >1% in cases or controls) and did not deviate from Hardy–Weinberg equilibrium in the control population ($P_{HWE} > 0.001$).

Mutation detection

Amplicons were generated using the following touchdown PCR program: 95 °C for 12 min, (95 °C for 30 s, T_m for 30 s, 72 °C for 30 s) × 16 cycles (td −0.5 °C), (95 °C for 30 s, T_m for 30 s, 72 °C for 30 s) × 19 cycles, 72 °C for 10 min and 10 °C for ∞.

Sequencing of genomic DNA was performed using BigDye Terminator v3.1 chemistry (Applied Biosystems) and an ABI3730 capillary sequencer (Applied Biosystems) according to manufacturer protocols. Traces were inspected for SNPs and InDels using novoSNP v2.03.

Statistical analyses

Our screening panel A had 80% power to detect a variant with an odds ratio of 1.48 or higher at the 5% significance level, assuming a frequency of the disease-associated allele of at least 30% in controls (calculated with PS Power and Sample Size v3.0.12[15]). For subsequent statistical analyses, we used a data set that passed stringent quality control filters (see section SNP genotyping for genome-wide screening) resulting in 355 750 SNPs genotyped in 419 cases and 474 healthy controls. The total genotyping rate across these samples was 99.8%. We found minimal evidence for an overall inflation of the test statistics due to population stratification with a moderate genomic control value of $\lambda_{GC} = 1.14$. Furthermore, a multi-dimensional scaling analysis showed genuine European ancestry of panel A and identity-by-state analysis revealed neither non-European 'outliers' nor cryptically related individuals after quality control (Supplementary Figure S1). Genome-wide association analyses were conducted with gPLINK v2.050 in combination with PLINK v1.07.[16] GWAS data were also analyzed with R statistical environment version 2.10.0.

After the initial comparison of case–control frequencies (Supplementary Figure S2), we pruned the SNP list for redundancy due to linkage disequilibrium (using the 'clump' command in PLINK) and visually inspected the cluster plots of the obtained index SNPs with a P-value $<2 \times 10^{-4}$ (Supplementary Figure S3). The 100 most strongly associated SNPs were subsequently selected for replication analysis and genotyped in the German replication panel B (consisting of 406 cases and 1682 controls) and the Italian replication panel C (comprising 287 cases and 579 controls). After quality control, we excluded 12 follow-up SNPs with a call rate less than 95% in cases or controls and one SNP with significant deviation from Hardy–Weinberg equilibrium in the controls ($P < 0.001$). To take the different geographic origin of the two replication panels into account, we used a Cochran–Mantel–Haenszel test (P_{CMH}) and a Breslow–Day test for odds ratio heterogeneity (P_{BD}) in both the combined analysis of replication panels B and C and in the overall study sample (panels A through C) (Supplementary Table S2). The eight SNPs with a P-value $<5 \times 10^{-7}$ in the combined analysis of panels A and B or A through C were additionally genotyped in 258 ETV6–RUNX1-positive ALL cases and 800 healthy controls from Germany (extended replication panel B: 664 cases and 2482 controls).

Imputation

The software package BEAGLE v3.1.1 [(ref. 17)] was used to impute the genotypes of autosomal SNPs based on the 1000 Genomes data. As input for imputation, we used only SNPs on the Affymetrix platform that passed quality control. To take imputation uncertainty into account, association analysis between the phenotype and the dosage data (expected allele counts) was performed using PLINK's logistic regression framework for dosage data (with the 'dosage' command).

Plotting

Regional association plots were generated by using a modified version of deBakker's R script. SNPs are represented with their corresponding P-values (on $-\log_{10}$ scale) on the vertical axis as a function of physical position. Single-marker association analysis was performed using PLINK's standard case–control allelic test (with the 'assoc' command) and study SNPs as well as imputed SNPs were used as input. Physical positions on chromosomes are based on NCBI build 36. Estimated recombination rates from phased haplotypes in HapMap release 22 (build 36) were downloaded from the HapMap website, gene annotations were downloaded from the UCSC genome browser (using build 36 coordinates).

eQTL analysis

Expression data of TP63 (TP73L), PTPRJ and INTS10 (FLJ10569) in EBV-transformed lymphoblastoid cell lines from the HapMap samples was available from the GENEVAR project,[18] and expression in the $n = 60$ CEU parents was correlated to SNP genotypes, using linear regression in the web-based tool SNPEXP v1.1. (http://app3.titan.uio.no/biotools/tool.php?app=snpexp).

RESULTS

Associations with ETV6–RUNX1 subtype of childhood ALL

Novel significant disease associations that withstood Bonferroni correction (adjusted significance threshold $\alpha = 5.75 \times 10^{-4}$ (0.05/87)) 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 ($P_{CMH} = 4.87 \times 10^{-7}$, 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 ($P_{CMH} = 2.54 \times 10^{-4}$, 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 ($P_{CMH} = 8.94 \times 10^{-9}$) 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 $P_{CMH} = 1.00 \times 10^{-6}$.

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 INTS10 (Integrator complex subunit 10; $P = 1.05 \times 10^{-5}$). 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 4.95×10^{-7} in the combined analysis (Table1 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 (⩾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[20]), we were not able to identify any significant correlation – so-called eQTL effect – between the genotypes at the lead SNPs rs17505012, 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[18]). For OR8U8 no expression data were available. However, it is noteworthy that the SNP rs3942852, located in the first intron of the PTPRJ gene,

Table 1. Summary of association results in combined replication panels B and C

Chr	position (bp)	dbSNP ID	Nearby genes (relative position)	A1	A2	panel A – GWAS Germany ETV6-RUNX1-pos. 474 controls 419 cases (AF_A1 contr./AF_A1 case)	P_{CCA}	panel B_extended Germany ETV6-RUNX1-pos. 2482 controls 664 cases (AF_A1 contr./AF_A1 case)	P_{CCA}	panel A + B_ext Germany ETV6-RUNX1-pos. 2956 controls 1083 cases (AF_A1 contr./AF_A1 case)	P_{CCA}	panel C Italy ETV6-RUNX1-pos. 579 controls 287 cases (AF_A1 contr./AF_A1 case)	P_{CCA}	P_{OMH}	panel B_ext + C Replication 3061 controls 951 cases P_{BD}	OR (95% CI)	panel A + B_ext. + C Combined 3535 controls 1370 cases P_{CMH}	P_{BD}
3	190884470	rs17505102	TP63 (Within gene)	C	G	0.18/0.11	1.15×10^{-04}	0.14/0.09	$\mathbf{1.50\times10^{-06}}$	0.15/0.10	2.96×10^{-08}	0.08/0.06	0.11	$\mathbf{4.87\times10^{-07}}$	0.44	0.63 (0.52–0.75)	8.94×10^{-09}	0.57
11	55932247	rs1945213	OR8U8 (Within gene)	C	C	0.33/0.33	5.82×10^{-05}	0.31/0.32	$\mathbf{1.63\times10^{-07}}$	0.31/0.32	9.14×10^{-11}	0.21/0.24	0.32	$\mathbf{4.27\times10^{-05}}$	4.13×10^{-04}	0.78 (0.69–0.88)	3.89×10^{-08}	2.21×10^{-04}
11	48071665	rs3942852	PTPRJ (Within gene)	G	G	0.24/0.24	1.58×10^{-04}	0.23/0.21	$\mathbf{1.67\times10^{-04}}$	0.24/0.22	4.95×10^{-07}	0.18/0.22	0.43	$\mathbf{2.54\times10^{-04}}$	0.21	0.77 (0.68–0.89)	1.00×10^{-06}	0.14
14	75773104	rs7156960	C14orf118 (±50 kb)	C	T	0.52/0.41	9.16×10^{-06}	0.47/0.42	$\mathbf{2.54\times10^{-03}}$	0.48/0.42	3.74×10^{-06}	0.41/0.37	0.10	$\mathbf{6.05\times10^{-04}}$	0.89	0.83 (0.75–0.92)	1.09×10^{-06}	0.58
15	90458377	rs207954[a]	SLCO3A1 (Within gene)	T	C	0.29/0.28	1.24×10^{-04}	0.28/0.25	0.09	0.28/0.23	6.58×10^{-05}	0.26/0.20	$\mathbf{5.60\times10^{-03}}$	$\mathbf{2.45\times10^{-03}}$	0.22	0.80 (0.70–0.93)	1.35×10^{-06}	0.61
8	19695441	rs920590	INTS10 (±50 kb)	C	T	0.29/0.38	5.19×10^{-05}	0.31/0.38	$\mathbf{1.05\times10^{-05}}$	0.31/0.38	6.12×10^{-10}	0.45/0.43	0.36	$\mathbf{1.32\times10^{-03}}$	1.87×10^{-03}	1.19 (1.07–1.33)	2.22×10^{-06}	5.54×10^{-06}
4	59198483	rs282708[a]	–	A	G	0.39/0.49	3.91×10^{-05}	0.41/0.45	$\mathbf{0.03}$	0.40/0.47	5.41×10^{-06}	0.46/0.48	0.42	0.03	0.51	1.15 (1.01–1.30)	1.38×10^{-05}	0.12
19	13727957	rs2910225[a]	CCDC130 (Within gene)	C	C	0.42/0.32	1.02×10^{-05}	0.36/0.33	0.16	0.37/0.32	6.97×10^{-04}	0.30/0.24	$\mathbf{0.01}$	$\mathbf{9.83\times10^{-03}}$	0.23	0.84 (0.74–0.96)	3.23×10^{-05}	0.54
15	97081324	rs11857366[a]	IGF1R (Within gene)	A	G	0.37/0.46	5.98×10^{-05}	0.40/0.43	0.08	0.39/0.45	7.01×10^{-05}	0.42/0.45	0.16	$\mathbf{0.02}$	0.98	1.15 (1.02–1.30)	3.41×10^{-05}	0.45
5	10085335	rs7734914[a]	–	A	G	0.40/0.50	1.32×10^{-05}	0.44/0.45	0.42	0.43/0.48	6.09×10^{-04}	0.46/0.52	$\mathbf{0.03}$	$\mathbf{4.71\times10^{-02}}$	0.21	1.13 (1.00–1.28)	4.81×10^{-05}	0.82
6	77846527	rs7738636	–	A	C	0.26/0.18	9.84×10^{-05}	0.23/0.20	$\mathbf{0.02}$	0.24/0.20	7.44×10^{-05}	0.16/0.15	0.46	$\mathbf{0.02}$	0.67	0.85 (0.75–0.97)	8.54×10^{-05}	0.35
1	111060293	rs343604[a]	KCNA3 (±50 kb)	T	C	0.07/0.03	8.39×10^{-05}	0.09/0.06	$\mathbf{0.03}$	0.08/0.05	1.19×10^{-06}	0.08/0.09	0.67	0.14	0.08	0.84 (0.67–1.06)	9.25×10^{-05}	1.46×10^{-03}
8	109041474	rs630662[a]	RSPO2 (Within gene)	A	G	0.30/0.22	3.61×10^{-05}	0.28/0.23	$\mathbf{4.46\times10^{-03}}$	0.29/0.22	2.02×10^{-06}	0.23/0.25	0.45	0.06	0.02	0.87 (0.76–1.01)	1.36×10^{-04}	2.65×10^{-03}
13	38845574	rs7336133[a]	LHFP (Within gene)	A	G	0.38/0.29	3.53×10^{-05}	0.32/0.31	0.42	0.34/0.30	6.01×10^{-03}	0.34/0.28	$\mathbf{0.01}$	$\mathbf{0.03}$	0.13	0.86 (0.75–0.98)	2.71×10^{-04}	0.39
1	66461697	rs17423910	PDE4B (Within gene)	A	G	0.14/0.23	2.18×10^{-06}	0.17/0.20	$\mathbf{0.01}$	0.17/0.21	6.37×10^{-04}	0.23/0.21	0.43	0.11	0.05	1.11 (0.98–1.27)	3.07×10^{-04}	0.01
6	143700705	rs6901152[a]	AIG1 (Within gene)	A	T	0.46/0.36	3.19×10^{-05}	0.42/0.36	$\mathbf{4.63\times10^{-03}}$	0.43/0.36	4.71×10^{-06}	0.44/0.47	0.37	0.09	0.01	0.90 (0.79–1.02)	4.19×10^{-04}	2.12×10^{-03}
4	159663764	rs936094[a]	RXFP1 (Within gene)	T	C	0.15/0.23	9.42×10^{-06}	0.17/0.21	$\mathbf{0.01}$	0.17/0.22	1.43×10^{-06}	0.25/0.23	0.28	0.24	0.02	1.09 (0.94–1.27)	4.74×10^{-04}	6.48×10^{-04}
3	55782295	rs6445754	ERC2 (Within gene)	C	T	0.17/0.28	1.82×10^{-07}	0.23/0.26	$\mathbf{7.97\times10^{-03}}$	0.22/0.27	4.60×10^{-06}	0.32/0.29	0.24	0.12	0.02	1.10 (0.98–1.24)	5.08×10^{-04}	1.62×10^{-03}
14	58908882	rs4901921[a]	DAAM1 (±10 kb)	T	C	0.52/0.43	1.48×10^{-04}	0.47/0.45	0.29	0.48/0.44	5.41×10^{-03}	0.40/0.35	$\mathbf{0.04}$	$\mathbf{0.04}$	0.32	0.88 (0.78–0.99)	6.45×10^{-04}	0.66
6	73500100	rs6910780[a]	KCNQ5 (Within gene)	C	T	0.16/0.23	1.47×10^{-04}	0.19/0.23	$\mathbf{0.01}$	0.18/0.23	3.89×10^{-04}	0.22/0.21	0.64	0.09	0.06	1.13 (0.98–1.31)	9.29×10^{-04}	0.02

Abbreviations: AF, allele frequency; Chr, chromosome; CI, confidence interval; GWAS, genome-wide association study; OR, odds ratio; SNP, single-nucleotide polymorphism. [a]The 12 SNPs that were not genotyped in the additional samples but only in the non-extended panel B: 1682 controls and 406 cases; panels B and C: 2156 controls and 825 cases; panels A and B: 2261 controls and 693 cases; panels A, B and C: 2735 controls and 1112 cases. 100 SNPs were selected for genotyping in two independent replication panels of ETV6-RUNX1-positive cases and healthy controls from Germany/Austria and Italy, respectively. (panels B and C). Data are shown for the 20 SNPs that were nominal significant either in replication panel B ($P_{CCA} < 0.05$) or in the combined replication analysis of panels B and C ($P_{CMH} < 0.05$ and $P_{BD} > 0.05$). Additionally, these 20 SNPs were also genotyped in 326 German ETV6-RUNX1-negative ALL cases (panel D) (Supplementary Table S3). The eight SNPs with a P-value $< 5\times10^{-7}$ in the combined analysis of panels A and B or A through C were also genotyped in 258 additional ETV6-RUNX1-positive ALL cases and 800 controls from Germany (panel B_extended 664 cases and 2482 controls). Results of all 100 SNPs are shown in Supplementary Table S2. SNPs are ranked according to the P-value obtained in the combined analysis of panels A, B and C. Nucleotide positions (position (bp)) refer to NCBI's build 36. A1 denotes the rare allele while A2 is the common allele. The respective allele frequencies are shown for allele A1 in controls and cases, respectively (AF_A1 contr./ AF_A1 case). P-values obtained in the case–control analysis using an allele based χ^2-test (one degree of freedom) are listed (P_{CCA}). Odds ratios (OR) and 95% confidence intervals (95% CI) for carriership of the allele A1 are shown for the replication analysis (panels B and C). Column P_{BD} lists the asymptotic P-values of a Breslow–Day test for heterogeneity. A significant P-value indicates a significant heterogeneity between replication panels in terms of the odds ratio of the disease association. Combined P-values (P_{CMH}) of the Cochran–Mantel–Haenszel test statistic (one degree of freedom) are shown. Significant P-values ($P_{CCA} < 0.05$ or $P_{CMH} < 0.05$ [only if $P_{BD} > 0.05$]) of the replication are highlighted in bold. SNP associations that remained significant after correction for multiple testing (P_{CMH} of panels B and C) as well as genome-wide significantly associated SNPs in the combined analysis (combined P-value of panels A and B or panels A–C) are highlighted by grey shading.
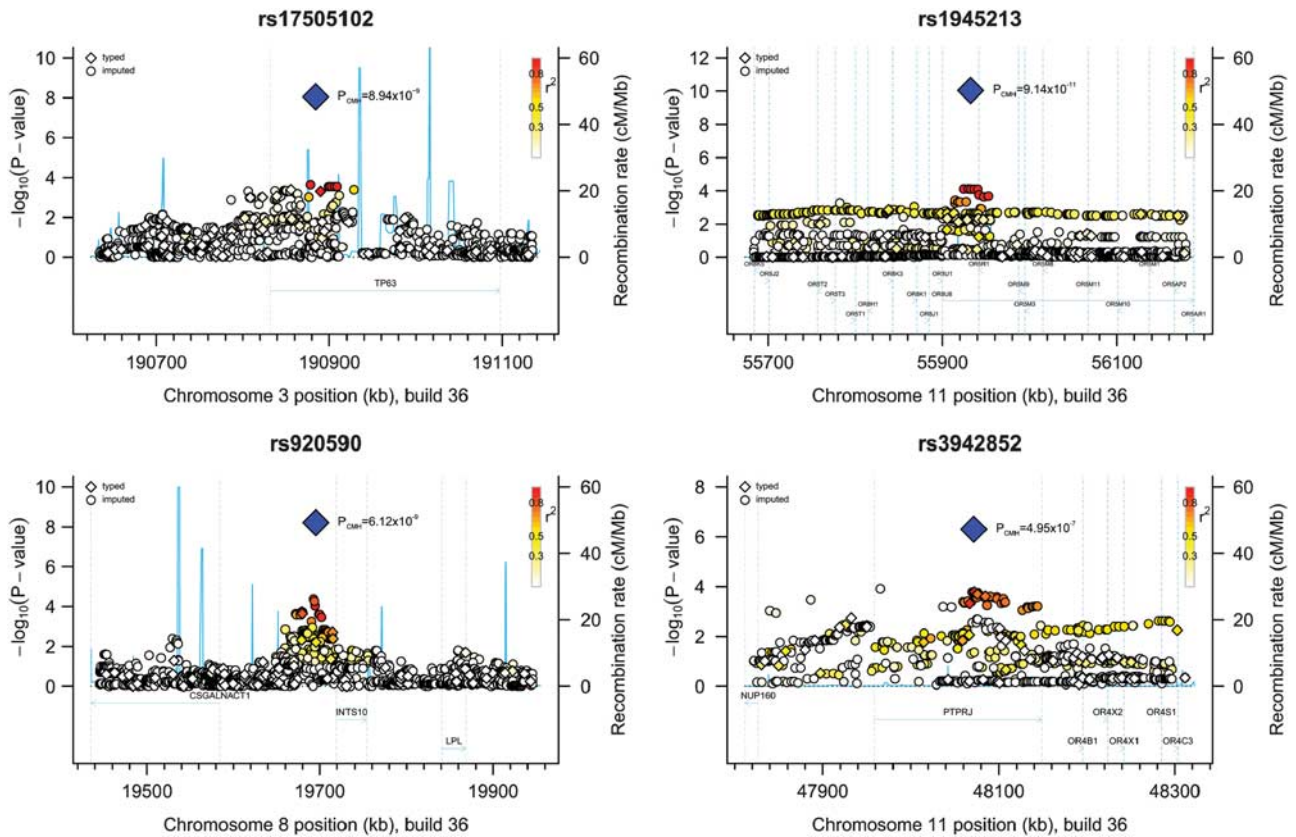
Figure 1. Regional plots of the confirmed associations at *TP63, OR8U8,* on 8p21.3 and at *PTPRJ*. The regional plots of the negative decadic logarithm of the *P*-values obtained in the GWAS (panel A) are shown. Panel A was imputed with CEU haplotypes generated by the 1000 Genomes Project (August 2010 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.

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 *ETV6–RUNX1* subtype

In order to assess whether our findings were specifically associated with the *ETV6–RUNX1*-rearranged subtype, we genotyped the 20 top SNPs displayed in Table 1 in a second German replication cohort (panel D) consisting of 326 *ETV6–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/B$_{extended}$ through D yielded only a marginally stronger association with $P_{CMH} = 1.42 \times 10^{-3}$ for rs7738636 (Supplementary Table S3). These results imply that rs7738636 might represent a susceptibility locus for *ETV6–RUNX1*-negative ALL patients. However, due to the small sample of *ETV6–RUNX1*-negative cases, our analysis does not have enough power to determine whether the newly identified loci distinguish *ETV6–RUNX1*-positive from -negative patients.

Associations of previously reported susceptibility-conferring SNPs

Additionally, we genotyped the susceptibility-conferring SNPs identified by Papaemmanuil *et al.*[11] and Treviño *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 *SIAT7C*, 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 *ETV6–RUNX1*-rearranged childhood ALL. Given their interesting biological function, we below discuss the relevance of *TP63* and *PTPRJ* in ALL disease etiology.

The *TP63* gene – a member of the *TP53* gene family – codes for p63 which has essential roles in embryonic development. *TP63* contains two transcriptional start sites leading to p63 isoforms either containing (TAp63) or lacking (△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

Table 2. Association results for known loci

Chr position (bp)	dbSNP ID / Gene name	A1	A2	Panel A Germany ETV6-RUNX1 pos. 474 controls 419 cases AF_{A1} contr.	AF_{A1} case	Panel B Germany ETV6-RUNX1 pos. 1682 controls 406 cases AF_{A1} contr.	AF_{A1} case	Panel C Italy ETV6-RUNX1 pos. 579 controls 287 cases AF_{A1} contr.	AF_{A1} case	Panel A+B+C Combined ETV6-RUNX1 pos. 2735 controls 1112 cases P_{CMH}	P_{BD}	OR (95% CI)	Panel D Germany ETV6-RUNX1 neg. — 326 cases AF_{A1} case	Panel A+B+C+D Combined ETV6-RUNX1 pos.+neg. 2735 controls 1438 cases P_{CMH}	P_{BD}	OR (95% CI)
Papaemmanuil et al. (2009)																
7 50 427 590	rs6964823 IKZF1	A	G	0.47	0.43	0.49	0.44	0.51	0.47	1.14×10^{-04}	0.85	0.82 (0.74–0.91)	0.44	5.56×10^{-05}	0.93	0.83 (0.75–0.91)
7 50 438 098	rs4132601 IKZF1	T	G	0.28	0.34	0.27	0.35	0.26	0.34	1.65×10^{-10}	0.58	1.42 (1.27–1.58)	0.35	8.26×10^{-13}	0.65	1.43 (1.30–1.58)
7 50 441 245	rs6944602 IKZF1	A	G	0.22	0.26	0.19	0.27	0.22	0.27	1.99×10^{-07}	0.39	1.42 (1.25–1.63)	0.26	3.10×10^{-09}	0.42	1.42 (1.27–1.60)
7 50 536 229	rs3779084 DDC	C	T	0.21	0.25	0.20	0.23	0.16	0.22	2.36×10^{-05}	0.17	1.30 (1.15–1.46)	0.25	2.44×10^{-06}	0.19	1.30 (1.17–1.46)
7 50 537 630	rs880028 DDC	C	T	0.21	0.25	0.20	0.23	0.16	0.22	2.91×10^{-05}	0.09	1.29 (1.15–1.46)	0.25	8.69×10^{-06}	0.09	1.28 (1.15–1.43)
7 50 540 827	rs7809758 DDC	T	A	0.25	0.35	0.23	0.34	0.22	0.32	2.04×10^{-05}	0.27	1.29 (1.15–1.45)	0.24	1.48×10^{-07}	0.40	1.32 (1.19–1.46)
10 63 369 901	rs7073837 ARID5B	A	C	0.40	0.42	0.37	0.44	0.43	0.49	7.76×10^{-07}	0.85	1.29 (1.17–1.42)	0.41	1.50×10^{-08}	0.75	1.30 (1.19–1.43)
10 63 388 485	rs10740055 ARID5B	A	C	0.51	0.47	0.53	0.46	0.49	0.42	8.75×10^{-07}	0.73	0.78 (0.70–0.86)	0.46	8.95×10^{-09}	0.88	0.76 (0.70–0.84)
10 63 422 165	rs7089424 ARID5B	G	T	0.33	0.38	0.32	0.42	0.38	0.44	9.05×10^{-10}	0.30	1.38 (1.24–1.53)	0.44	1.98×10^{-13}	0.17	1.42 (1.30–1.56)
14 22 658 897	rs2239633 CEBPE	T	C	0.48	0.42	0.50	0.41	0.43	0.38	1.53×10^{-08}	0.51	0.75 (0.67–0.83)	0.41	4.00×10^{-10}	0.50	0.74 (0.68–0.82)
Trevino et al. (2009)																
1 76 544 916	rs10873876 SIAT7C	T	C	0.17	0.15	0.14	0.19	0.18	0.22	8.14×10^{-03}	0.74	1.20 (1.05–1.36)	0.17	5.47×10^{-03}	0.71	1.19 (1.05–1.34)
1 195 111 216	rs6428370 –	C	G	0.30	0.31	0.31	0.32	0.36	0.40	0.14	0.41	1.08 (0.97–1.20)	0.35	0.03	0.63	1.12 (1.01–1.23)
1 235 333 226	rs7554607 RYR2	A	G	0.45	0.48	0.43	0.41	0.46	0.51	0.12	0.32	1.08 (0.98–1.20)	0.35	0.34	0.18	1.05 (0.95–1.15)
2 245 756 155	rs1881797 OR2C3	G	A	0.17	0.17	0.19	0.19	0.17	0.19	0.71	0.42	1.03 (0.90–1.17)	0.41	0.76	0.25	0.98 (0.87–1.11)
2 236 626 227	rs12621643 –	A	T	0.29	0.31	0.28	0.34	0.42	0.37	0.14	2.16×10^{-03}	1.08 (0.97–1.20)	0.16	0.06	1.81×10^{-03}	1.10 (1.00–1.21)
3 179 912 633	rs9290663 KCNE4	T	G	0.15	0.16	0.14	0.17	0.23	0.27	0.02	0.70	1.18 (1.03–1.34)	0.32	6.90×10^{-03}	0.74	1.18 (1.05–1.33)
6 141 211 518	rs11155133 KCNMB2	G	A	0.01	0.01	0.01	0.01	0.01	0.01	0.27	0.37	1.32 (0.81–2.14)	0.17	0.28	0.41	1.28 (0.81–2.03)
7 50 433 798	rs11978267 IKZF1	G	A	0.29	0.35	0.27	0.34	0.25	0.33	8.74×10^{-11}	0.70	1.43 (1.28–1.59)	0.01	1.07×10^{-12}	0.73	1.44 (1.30–1.59)
7 50 496 943	rs2242041 DDC	A	C	0.09	0.10	0.09	0.12	0.05	0.08	7.51×10^{-03}	0.38	1.26 (1.06–1.50)	0.35	8.47×10^{-04}	0.48	1.30 (1.11–1.52)
7 50 533 321	rs2167364 DDC	C	A	0.33	0.37	0.31	0.37	0.29	0.37	2.33×10^{-07}	0.35	1.32 (1.19–1.46)	0.12	1.73×10^{-08}	0.36	1.32 (1.20–1.45)
10 34 857 994	rs563507 PARD3	A	G	0.04	0.06	0.04	0.04	0.07	0.08	0.047	0.65	1.25 (1.00–1.55)	0.37	0.046	0.73	1.23 (1.00–1.51)
10 63 380 110	rs10994982 ARID5B	A	G	0.51	0.52	0.46	0.52	0.51	0.57	2.07×10^{-05}	0.86	1.24 (1.12–1.37)	0.05	1.39×10^{-06}	0.94	1.25 (1.14–1.37)
10 63 393 583	rs10821936 ARID5B	C	T	0.33	0.38	0.31	0.42	0.37	0.44	1.25×10^{-11}	0.56	1.42 (1.29–1.58)	0.53	4.06×10^{-15}	0.42	1.46 (1.33–1.60)
12 4 295 383	rs10849033 C12orf5	G	A	0.02	0.01	0.01	0.03	0.02	0.03	0.01	0.69	1.62 (1.11–2.37)	0.42	0.046	0.91	1.43 (1.00–2.02)
12 115 487 041	rs2089222 KRTHB5	A	G	0.03	0.04	0.04	0.04	0.03	0.03	0.58	0.52	1.08 (0.83–1.41)	0.03	0.87	0.40	1.02 (0.80–1.31)

Table 2 (Continued)

Chr position (bp)	dbSNP ID / Gene name	A1 / A2	Panel A Germany ETV6-RUNX1 pos. 419 cases / 474 controls AF_A1 contr. / AF_A1 case	Panel B Germany ETV6-RUNX1 pos. 1682 controls / 406 cases AF_A1 contr. / AF_A1 case	Panel C Italy ETV6-RUNX1 pos. 579 controls / 287 cases AF_A1 contr. / AF_A1 case	Panel A+B+C Combined analysis ETV6-RUNX1 pos. 2735 controls / 1112 cases P_{CMH}	P_{BD}	OR (95% CI)	Panel D Germany ETV6-RUNX1 neg. 326 cases AF_A1 case	Panel A+B+C+D Combined analysis ETV6-RUNX1 pos. + neg. 2735 controls / 1438 cases P_{CMH}	P_{BD}	OR (95% CI)
18 2488054	rs1879352	G / A	0.15 / 0.14	0.16 / 0.14	0.17 / 0.19	0.75	0.09	0.98 (0.85–1.12)	0.15	0.83	0.11	0.99 (0.87–1.12)
19 49203229	rs2191566 ZNF230	C / A	0.29 / 0.29	0.29 / 0.30	0.34 / 0.29	0.35	0.04	0.95 (0.85–1.06)	0.33	0.88	0.01	0.99 (0.90–1.10)
19 49214753	rs6509133 ZNF230	C / T	0.29 / 0.29	0.30 / 0.32	0.35 / 0.30	0.46	0.04	0.96 (0.86–1.07)	0.33	0.98	0.02	1.00 (0.91–1.10)

Abbreviations: AF, allele frequency; Chr, chromosome; CI, confidence interval; OR, odds ratio; SNP, single-nucleotide polymorphism. In order to confirm the associated SNPs reported by Papaemmanuil et al. and Treviño et al., we genotyped the respective 28 SNPs in the panels A through D. SNPs are ranked according to their chromosomal position and to the respective study. Nucleotide positions (position (bp)) refer to NCBI's build 36. The rare allele is A1 while A2 denotes the common allele. The respective allele frequencies are listed for allele A1 in controls and cases, respectively, (AF_A1 contr./AF_A1 case). Combined P-values (P_{CMH}) and combined odds ratios (OR) of the Cochran–Mantel–Haenszel test statistic (one degree of freedom) for the case–control analysis of panels A through C and panels A through D are listed. Column P_{BD} shows the asymptotic P-values of a Breslow–Day test for heterogeneity. A significant P_{BD} indicates a significant heterogeneity between replication panels in terms of the odds ratio of the disease association. Significant P-values ($P_{CMH} < 0.05$ [only if $P_{BD} > 0.05$]) are highlighted in bold. SNP associations that reached genome-wide significance in the combined panels A through C or A through D are highlighted by grey shading.

rarely mutated in human cancers.[27,28] To date, recurrent somatic *TP63* mutations have only been described for non-small cell lung cancer as well as chronic myeloid leukemia[28,29] and *TP63* germline genetic variation (rs710521) has only been associated with urinary bladder cancer, so far.[30] 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.[31] 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-dependant manner.[32,33] 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.[33] Also, it was found that thymomas and certain B-cell lymphomas express high levels of TP63, suggesting a role of TP63 in these cancers.[34]

PTPRJ 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 *PTPRJ* have been described in a wide spectrum of cancers like thyroid carcinomas, colon cancer, breast and lung cancer.[35-38] Furthermore, a specific *PTPRJ* haplotype was shown to be associated with the risk of breast cancer demonstrating that common *PTPRJ* variants may act as breast cancer susceptibility alleles.[39] Interestingly, the *PTPRJ* gene is expressed at high levels in hematopoietic cells.[35,37,40] Of particular interest to the results reported in our study, mice with a constitutively deleted transmembrane domain allele of *PTPRJ* display a partial peripheral B-cell developmental block.[41] Heterozygous mice do not show any notable phenotype.

Although the current knowledge on *TP63* and *PTPRJ* 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 *PTPRJ* loci may modulate susceptibility to precursour 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 *IKZF1*, *DDC*, *ARID5B* and *CEBPE*. Nominal significant P-values ($P < 0.05$) were obtained for *SIAT7C*, 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.[11] and Trevino et al.,[12] respectively, we could not replicate the loci *RYR2*, *OR2C3*, *KCNE4*, 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.

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, RH, AN and MZ helped with data analysis. PN coordinated the Affymetrix chip genotyping. M St, 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

SNPexp[20] v1.1:

http://app3.titan.uio.no/biotools/tool.php?app = snpexp

R script for Regional Association Plots:

http://www.broadinstitute.org/science/projects/diabetes-genetics-initiative/plotting-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, Wilimas JA, Quintana Y, Pui CH, Robison LL et al. Childhood cancer epidemiology in low-income countries. *Cancer* 2008; **112**: 461–472.

6 Belson 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-Wasserman Lecture. *Hematol Am Soc Hematol Educ Program* 2009; **2009**: 3–12, PMID: 20008176.

11 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**: 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, 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**: 1765–1767.

14 Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, 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.

19 Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001; **29**: 308–311.

20 Holm K, Melum E, Franke A, Karlsen TH. SNPexp - a web tool for calculating and visualizing correlation between HapMap genotypes and gene expression levels. *BMC Bioinformatics* 2010; **11**: 600.

21 Collavin L, Lunardi A, Del Sal G. p53-family proteins and their regulators: hubs and spokes in tumor suppression. *Cell Death Differ* 2010; **17**: 901–911.

22 Mills AA, Zheng B, Wang XJ, Vogel H, Roop DR, Bradley A. p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature* 1999; **398**: 708–713.

23 Yang A, Schweitzer R, Sun D, Kaghad M, Walker N, Bronson RT et al. p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* 1999; **398**: 714–718.

24 Senoo M, Pinto F, Crum CP, McKeon F. p63 Is essential for the proliferative potential of stem cells in stratified epithelia. *Cell* 2007; **129**: 523–536.

25 Su X, Paris M, Gi YJ, Tsai KY, Cho MS, Lin YL et al. TAp63 prevents premature aging by promoting adult stem cell maintenance. *Cell Stem Cell* 2009; **5**: 64–75.

26 Celli J, Duijf P, Hamel BC, Bamshad M, Kramer B, Smits AP et al. Heterozygous germline mutations in the p53 homolog p63 are the cause of EEC syndrome. *Cell* 1999; **99**: 143–153.

27 Hagiwara K, McMenamin MG, Miura K, Harris CC. Mutational analysis of the p63/p73L/p51/p40/CUSP/KET gene in human cancer cell lines using intronic primers. *Cancer Res* 1999; **59**: 4165–4169.

28 Sunahara M, Shishikura T, Takahashi M, Todo S, Yamamoto N, Kimura H et al. Mutational analysis of p51A/TAp63gamma, a p53 homolog, in non-small cell lung cancer and breast cancer. *Oncogene* 1999; **18**: 3761–3765.

29 Yamaguchi H, Inokuchi K, Sakuma Y, Dan K. Mutation of the p51/p63 gene is associated with blastic crisis in chronic myelogenous leukemia. *Leukemia* 2001; **15**: 1729–1734.

30 Kiemeney LA, Sulem P, Besenbacher S, Vermeulen SH, Sigurdsson A, Thorleifsson G et al. A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. *Nat Genet* 2010; **42**: 415–419.

31 Flores ER, Sengupta S, Miller JB, Newman JJ, Bronson R, Crowley D et al. Tumor predisposition in mice mutant for p63 and p73: evidence for broader tumor suppressor functions for the p53 family. *Cancer Cell* 2005; **7**: 363–373.

32 Lantner F, Starlets D, Gore Y, Flaishon L, Yamit-Hezi A, Dikstein R et al. CD74 induces TAp63 expression leading to B-cell survival. *Blood* 2007; **110**: 4303–4311.

33 Binsky I, Lantner F, Grabovsky V, Harpaz N, Shvidel L, Berrebi A et al. TAp63 regulates VLA-4 expression and chronic lymphocytic leukemia cell migration to the bone marrow in a CD74-dependent manner. *J Immunol* 2010; **184**: 4761–4769.

34 Di Como CJ, Urist MJ, Babayan I, Drobnjak M, Hedvat CV, Teruya-Feldstein J et al. p63 expression profiles in human normal and tumor tissues. *Clin Cancer Res* 2002; **8**: 494–501.

35 Mita Y, Yasuda Y, Sakai A, Yamamoto H, Toyooka S, Gunduz M et al. Missense polymorphisms of PTPRJ and PTPN13 genes affect susceptibility to a variety of human cancers. *J Cancer Res Clin Oncol* 2010; **136**: 249–259.

36 Iuliano R, Le Pera I, Cristofaro C, Baudi F, Arturi F, Pallante P et al. The tyrosine phosphatase PTPRJ/DEP-1 genotype affects thyroid carcinogenesis. *Oncogene* 2004; **23**: 8432–8438.

37 Ruivenkamp CA, van Wezel T, Zanon C, Stassen AP, Vlcek C, Csikos T et al. Ptprj is a candidate for the mouse colon-cancer susceptibility locus Scc1 and is frequently deleted in human cancers. *Nat Genet* 2002; **31**: 295–300.

38 Ruivenkamp C, Hermsen M, Postma C, Klous A, Baak J, Meijer G et al. LOH of PTPRJ occurs early in colorectal cancer and is associated with chromosomal loss of 18q12-21. *Oncogene* 2003; **22**: 3472–3474.

39 Lesueur F, Pharoah PD, Laing S, Ahmed S, Jordan C, Smith PL et al. Allelic association of the human homologue of the mouse modifier Ptprj with breast cancer. *Hum Mol Genet* 2005; **14**: 2349–2356.

40 de la Fuente-Garcia MA, Nicolas JM, Freed JH, Palou E, Thomas AP, Vilella R et al. CD148 is a membrane protein tyrosine phosphatase present in all hematopoietic lineages and is involved in signal transduction on lymphocytes. *Blood* 1998; **91**: 2800–2809.

41 Zhu JW, Brdicka T, Katsumoto TR, Lin J, Weiss A. Structurally distinct phosphatases CD45 and CD148 both regulate B cell and macrophage immunoreceptor signaling. *Immunity* 2008; **28**: 183–196.

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