Published in final edited form as:
Nat Genet. 2009 September ; 41(9): 1006–1010. doi:10.1038/ng.430.

Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia

Elli Papaemmanuil[1], **Fay J Hosking**[1], **Jayaram Vijayakrishnan**[1], **Amy Price**[1], **Bianca Olver**[1], **Eammon Sheridan**[2], **Sally E Kinsey**[3], **Tracy Lightfoot**[4], **Eve Roman**[4], **Julie A E Irving**[5], **James M. Allan**[5], **Ian P Tomlinson**[6], **Malcolm Taylor**[7], **Mel Greaves**[8], and **Richard S Houlston**[1,*]

[1]Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK

[2]Yorkshire Regional Genetic Service, St James's University Hospital, Beckett Street, Leeds, LS9 7T, UK

[3]Department of Paediatric and Adolescent Oncology and Haematology, St James University Hospital, Leeds, LS9 7T, UK

[4]Epidemiology and Genetics Unit, Department of Health Sciences, University of York, York Y010 5DD, UK

[5]Northern Institute for Cancer Research, Newcastle University, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK

[6]Molecular and Population Genetics, Wellcome Trust Centre for Human Genetics, Oxford, OX3 7BN, UK

[7]Cancer Immunogenetics Group, School of Cancer Sciences, University of Manchester, Research Floor, St Mary's Hospital, Manchester, M13 0JH, UK

[8]Section of Haemato-oncology, Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK

Abstract

To identify risk variants for childhood acute lymphoblastic leukemia (ALL) we conducted a genome-wide association study of 2 case-control series, analyzing the genotypes of 291,423 tagging SNP genotypes in a total of 907 ALL cases and 2,398 controls. We identified risk loci for ALL at 7p12.2 (*IKZF1*, rs4132601; OR = 1.69, $P = 1.20 \times 10^{-19}$), 10q21.2 (*ARIDB5*, rs7089424;

*Corresponding author: Richard.houlston@icr.ac.uk, Tel: +44 (0) 208 722 4175, Fax: +44 (0) 208 722 4365.

Author Contributions

RSH and MG designed the study and obtained financial support. RSH drafted the manuscript with, substantial contributions from MG. EP performed overall project management, development, database development and oversaw laboratory analyses; FJH performed statistical analyses; FJH and EP performed bioinformatics analyses; JV, BO, AP performed sample preparation; ES and SEK performed curation and sample preparation of MRC ALL 97 trial samples; TL and ER management and maintenance of UKCCS sample data; MT performed curation and sample preparation of UKCCS samples; JMA and JAEI performed ascertainment, curation and sample preparation of Northern Institute for Cancer Research case series. IPT performed generation and management of UK CRC control genotypes. All authors contributed to the final paper.

Competing Interests Statement

The authors declare no competing financial interests

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

OR = 1.65, $P = 6.69 \times 10^{-19}$) and 14q11.2 (*CEBPE*, rs2239633; OR = 1.34, $P = 2.88 \times 10^{-7}$). The 10q21.2 (*ARIDB5*) risk association appears to be selective for the subset of B-cell precursor ALL with hyperdiploidy. These data show that common low-penetrance susceptibility alleles contribute to the risk of developing childhood ALL and provide novel insight into disease causation of this hematological cancer; notably all 3 risk variants map to genes involved in transcriptional regulation and differentiation of B-cell progenitors.

Acute leukemia is the major pediatric cancer in developed countries, where it affects 30-45 per 10^6 children per year1. The childhood acute leukemias are heterogeneous with respect to their underlying cellular and molecular biology, acquired genetic abnormalities and associated clinical responses to combination chemotherapy2. It is suspected therefore that the acute myeloblastic leukemia and subtypes of B or T cell precursor acute lymphoblastic leukemia (ALL) may not share a common etiology3.

While epidemiology data are compatible with transplacental carcinogen exposure as a basis for infant leukemias associated with *MLL* gene fusion3 and dysregulated immune response to common infection is a candidate for childhood ALL3, the role of environmental carcinogenesis in childhood leukemia is presently undefined. It is however probable that the risk of ALL from environmental exposure is influenced by genetic variation. Data from the Swedish family-cancer database lends support to a small familial risk of ALL4,5, independent of the high concordance in monozygotic twins (which has a non-genetic, *in-utero* explanation). Although rare (<5% of ALL) direct evidence for an inherited genetic predisposition to ALL is provided by the high risk associated with Bloom's syndrome, neurofibromatosis, ataxia telangiectasia and constitutional trisomy 216. The heritable basis of susceptibility to ALL outside these syndromes is presently undefined but it is likely that the co-inheritance of multiple low-risk variants contribute to disease risk. Predicated on this hypothesis we have conducted a genome-wide association (GWA) study of ALL analyzing 2 case series genotyped using Illumina Infinium HD Human370 Duo BeadChips.

The first GWA study (GWA-1) was based on genotyping 577 ALL cases from the United Kingdom Childhood Cancer Study (UKCCS)7. After applying quality control criteria and exclusion of individuals with non-Western European genotype, SNP genotype data were available for 503 of the cases (Supplementary Fig. 1). For controls we made use of publicly accessible Illumina Hap550K BeadChip genotype data generated on 1,438 individuals from the British 1958 Birth Cohort (58C, also known as the National Child Development Study), which included all live births in England, Wales and Scotland during a single week in 19588.

The second GWA study (GWA-2) was based on genotyping 392 cases from the Medical Research Council (MRC) ALL 97 trial and 36 cases collected by the Northern Institute for Cancer Research (NICR). After applying quality control criteria and exclusion of individuals with non-Western European genotype, SNP genotype data were available for 404 of the cases (Supplementary Fig. 1). For controls we made use of Illumina Hap550K BeadChip genotype data generated on 960 healthy individuals from the UK as part of a study of colorectal cancer (CRC) 9.

Across both case series a total of 342,665 tagging SNPs were satisfactorily genotyped (99.7%), with mean individual sample call rates (the percentage of samples for which a genotype was obtained for each SNP) of 99.8%. We excluded 24 individuals because of non-Western European ancestry and 10 because of cryptic relatedness (Supplementary Fig. 2). 293,371 SNP genotypes were available on all 907 cases (824 B-cell ALL, 83 T-cell ALL) and 2,398 controls in the combined data (Supplementary Fig. 1). Prior to undertaking a meta-analysis we searched for potential errors and biases in the 2 GWA studies imposing a high stringency for quality control for SNPs. Specifically, we considered only the 291,473 autosomal SNPs which had call rates >95% in all case and control series, which showed no extreme departure from Hardy-Weinberg equilibrium (HWE; $P>10^{-5}$ in controls) and had minor allele frequencies (MAF) exceeding 1% in cases and controls (Supplementary Fig. 1).

Comparison of the observed and expected distributions showed little evidence for an inflation of the test statistics in the 2 datasets (inflation factor10 $\lambda = 1.034$ and 1.002 for GWA-1 and GWA-2 respectively based on the 90% least significant SNPs; Supplementary Fig. 3), thereby excluding the possibility of significant hidden population substructure or differential genotype calling between cases and controls. Using data from both GWA studies we derived joint odds ratios (ORs) and confidence intervals (CI) under a fixed effects model for each SNP and associated *P* values from the standard normal distribution.

Ten SNPs mapping to 3 genomic regions showed evidence of an association at conventional levels for genome-wide significance (i.e. $P < 5 \times 10^{-7}$)11. These associations were significant at this threshold irrespective of which control group was used for reference (Supplementary Table 1).

The strongest association signal was attained at 7p12.2 with rs4132601 (combined OR = 1.69, 95% CI: 1.58 – 1.81; $P = 1.20 \times 10^{-19}$; $P_{het} = 0.68$, $I^2 = 0\%$; Table 1, Supplementary Table 1), which maps to the 3' region of the ikaros family zinc finger 1 (*IKZF1*) gene (50,438,098 bps; Fig. 1). The association signal was also highly significant when analysis was confined to B-cell ALL (combined OR = 1.73, 95% CI: 1.61 – 1.85; $P = 9.31 \times 10^{-20}$; Table 1, Supplementary Table 1). rs6944602 and rs6964823 which map to the 3' UTR and intron 7 of *IKZF1* (50,441,245 bps and 50,427,590 bps; Fig. 1; Supplementary Fig. 4), also displayed statistical support for an association at 7p12.2 at genome-wide threshold ($P = 6.02 \times 10^{-14}$ and 3.43×10^{-15} respectively). rs4132601, rs6964823 and rs6944602 map to a 27.3kb block of linkage disequilibrium (LD). LD metrics for rs4132601/rs6964823, rs4132601/rs6944602, rs6964823/rs6944602 are $D' = 1.0, 1.0, 1.0$ and $r^2 = 0.42, 0.79, 0.33$ respectively. A significant association was also seen with rs7809758, which maps 110kb centromeric to rs4132601 (50,540,827 bps; $P = 2.41 \times 10^{-10}$), which annotates the Dopa decarboxylase aromatic L-amino acid (*DCC*) gene. LD between rs7809758 and rs4132601 is not strong ($D' = 0.72$, $r^2 = 0.32$) and the 9-order difference in statistical support for the association with ALL defined by rs7809758 compared with rs4132601 argues against an independent disease-locus. However, to confirm this will require further mapping and dissection of recombination breakpoints.

Although rs4132601 may not be directly functional the established role of *IKZF1* in the biology of ALL strongly implicates variation in *IKZF1* as the causal basis of the 7p12.2

association. Ikaros proteins are master regulators of lymphocyte development and differentiation plays a pivotal role in CD4 versus CD8 T-cell lineage commitment decisions12. Germline mutant mice expressing only non-DNA-binding dominant-negative leukemogenic Ikaros isoforms develop an aggressive form of lymphoblastic leukemia13,14. Chromosomal deletions involving IKZF1 are common (30%) in high-risk/poor prognosis B-cell precursor ALL15 and are highly prevalent (95%) in ALL with BCR-ABL1 fusions16.

To explore the possibility that the association might be mediated through differential *IKZF1* expression we investigated the relationship between rs4132601 genotype and mRNA expression level in EBV transformed lymphocytes. Expression was significantly associated with genotype in a dose-dependent fashion ($P = 0.005$; Fig. 2, Supplementary Fig. 5) with lower levels being associated with risk alleles. This observation is consistent with a model in which the causal variant influences risk by reducing the efficiency of early B-cell differentiation.

The second strongest association signal was attained at 10q21.2 with rs7089424 (combined OR = 1.65, 95% CI: 1.54 – 1.76; $P = 6.69 \times 10^{-19}$; $P_{het} = 0.29$, $I^2 = 7\%$; Table 1, Supplementary Table 1), which maps to intron 3 of the AT rich interactive domain 5B (*ARIDB5*) gene (63,422,165 bps; Fig. 1; Supplementary Fig. 4). The association signal was also highly significant when analysis was confined to B-cell ALL (combined OR = 1.70, 95% CI: 1.58 – 1.81; $P = 1.41 \times 10^{-19}$; Table 1, Supplementary Table 1). Two additional SNPs rs7073837 and rs10740055 having amongst the most extreme *P* values ($P = 4.66 \times 10^{-16}$ and 5.35×10^{-13} respectively) also annotate *ARIDB5* (63,369,901bps and 63,388,485bps), localizing to introns 2 and 3 of the gene respectively. These are in LD with rs7089424 ($D' = 0.89$, 1.0 and $r^2 = 0.60$, 0.43 respectively) and map to a 79.3kb block of LD within *ARIDB5*. ARIDB5 is member of the AT-rich interaction domain family of transcription factors17, which plays an important role in embryogenesis and growth regulation18. While *ARIDB5* expression is reported to be upregulated in acute promyelocytic leukemia19, currently there is no evidence for involvement of *ARIDB5* in childhood ALL. Evidence for *ARIDB5* having a role in defining B-cell lineage is supported by data from homozygous knockout mice which along with growth retardation phenotype have decreased bone marrow cellularity and reduced numbers of B-cell progenitors18.

The third strongest statistical evidence for an association was attained at 14q11.2 with rs2239633 (22,658,897 bps; combined OR = 1.34, 95% CI: 1.22 – 1.45; $P = 2.88 \times 10^{-7}$; $P_{het} = 0.22$, $I^2 = 33\%$; Table 1, Supplementary Table 1). As with the previous regions the association signal was also significant when analysis was confined to B-cell ALL (combined OR = 1.37, 95% CI: 1.26 – 1.49; $P = 5.60 \times 10^{-8}$; Table 1, Supplementary Table 1). rs2239633 maps within a 25.7kb region of LD which encompasses the gene encoding CCAAT/enhancer-binding protein, epsilon (*CEBPE*; Fig. 1; Supplementary Fig. 4). Two other SNPs associated with ALL risk at $P < 10^{-5}$ (rs7157021 and rs10143875) map within this region of LD, providing additional support for 14q11.2 as a susceptibility locus. CEBP is a suppressor of myeloid leukemogenesis and is mutated in a subset of cases. Intriguingly, *CEBPE*, along with other *CEBP* family members has been shown to be occasionally (~1%) targeted by recurrent IGH translocations in B-cell precursor ALL20 suggesting opposing

functions of CEBP dysregulation in myeloid and lymphoid leukemogenesis and a possible role in susceptibility to ALL.

Given the biological heterogeneity of ALL, we analyzed the association between the major subtypes of ALL and rs4132601, rs7089424 and rs2239633 genotypes through case-only logistic regression (Table 2). The primary impact of variation defined by the 7p12.2, 10q21.2 and 14q11.2 risk variants is for B-lineage leukemia. Subtype analysis of B precursor ALL provides strong evidence that variation at 10q21.2-*ARIDB5* is highly associated with the risk of developing hyperdiploid ALL ($P = 3.84 \times 10^{-6}$; Table 2).

Fine-mapping and resequencing is required to identify the specific functional variant underlying each of the associations we have identified. Accepting HapMap is not comprehensive; few non-synonymous SNPs have been documented in *IKZF1, ARID5B*, and *CEBPE* and none are correlated with rs4132601, rs7089424 or rs2239633. These data suggest the associations identified are mediated through LD with sequence changes that influence gene expression rather than protein sequence or through LD with low frequency variants that are not catalogued by HapMap. While we did not find a significant relationship between genotypes and *ARID5B* and *CEBPE* expression (Supplementary Fig. 5) we were able to demonstrate a relationship between rs4132601 genotype and *IKZF1* expression compatible with the causal variant at this locus influencing differential expression.

When we modeled pairwise combinations of the SNPs, we did not find evidence of interactive effects between any of the three loci identified ($P > 0.1$ for all pairwise interactions: Supplementary Table 2), suggesting that each locus has an independent role in ALL development. While the risks of ALL associated with 7p12.2, 10q21.2 and 14q11.2 variants are modest the carrier frequencies of risk alleles of rs4132601, rs7089424 and rs2239633 are high in the European population and hence the loci make a significant contribution to the development of ALL, underlying ~64% of cases.

Our findings provide the first unambiguous evidence that common genetic variation influences the risk of developing pediatric ALL and a strong rationale for searching for additional risk variants through additional GWA scans. Furthermore, these findings provide novel insight into the development of ALL. It is striking that the 3 risk variants we identify map to genes involved in transcriptional regulation and differentiation of B-cell progenitors. Ethnic differences in the risk of ALL are well recognized1, it will therefore be interesting to explore how our findings translate to non-Western European populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Leukemia Research (UK) and the Kay Kendall Leukemia Fund provided principal funding for the study. Support from Cancer Research UK (C1298/A8362 supported by the Bobby Moore Fund) is also acknowledged. The study made use of genotyping data on the 1958 Birth Cohort. Genotyping data on 1958 controls was generated and generously supplied to us by Panagiotis Deloukas of the Wellcome Trust Sanger Institute. A full list of the investigators who contributed to the generation of the 1958 data is available from www.wtccc.org.uk/. We acknowledge use of DNA from the British 1958 Birth Cohort collection, funded by the Medical Research Council

Europe PMC Funders Author Manuscripts

grant G0000934 and the Wellcome Trust grant 068545/Z/02. We are grateful to Sue Richards and Julie Burrett (Clinical Trials Service Unit, Oxford), Christine Harrison, Lucy Chilton and Anthony Moorman (Leukaemia Research Cytogenetics Group, Northern Institute for Cancer Research, Newcastle University), Jill Simpson (University of York), Pamela Thomson and Adiba Hussain (Cancer Immunogenetics, School of Cancer Sciences, University of Manchester) for assistance with data harmonization. Also to Irene Roberts and The Children's Cancer and Leukemia Group Biological Studies Steering Group for access to MRC ALL Trial samples. Pamela Thomson is funded by Children with Leukemia and we acknowledge their support. We acknowledge NHS funding to the NIHR Biomedical Research Centre. Finally, we are grateful to all the patients and individuals for their participation. We would also like to thank the clinicians and other hospital staff, and study staff who contributed to the blood sample and data collection for this study.

References

1. Stiller CA, Parkin DM. Geographic and ethnic variations in the incidence of childhood cancer. Br Med Bull. 1996; 52:682–703. [PubMed: 9039726]
2. Pui CH, Relling MV, Downing JR. Acute lymphoblastic leukemia. N Engl J Med. 2004; 350:1535–48. [PubMed: 15071128]
3. Greaves M. Infection, immune responses and the aetiology of childhood leukaemia. Nat Rev Cancer. 2006; 6:193–203. [PubMed: 16467884]
4. Hemminki K, Jiang Y. Risks among siblings and twins for childhood acute lymphoid leukaemia: results from the Swedish Family-Cancer Database. Leukemia. 2002; 16:297–8. [PubMed: 11840299]
5. Couto E, Chen B, Hemminki K. Association of childhood acute lymphoblastic leukaemia with cancers in family members. Br J Cancer. 2005; 93:1307–9. [PubMed: 16288301]
6. Hodgson, S.; Maher, E. A practical guide to human cancer genetics. Cambridge: Cambridge University Press; 1999. p. 116-122.
7. The United Kingdom Childhood Cancer Study: objectives, materials and methods. UK Childhood Cancer Study Investigators. Br J Cancer. 2000; 82:1073–102. [PubMed: 10737392]
8. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). Int J Epidemiol. 2006; 35:34–41. [PubMed: 16155052]
9. Tomlinson IP, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet. 2008; 40:623–30. [PubMed: 18372905]
10. Clayton DG, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet. 2005; 37:1243–6. [PubMed: 16228001]
11. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–78. [PubMed: 17554300]
12. Harker N, et al. The CD8alpha gene locus is regulated by the Ikaros family of proteins. Mol Cell. 2002; 10:1403–15. [PubMed: 12504015]
13. Georgopoulos K, et al. The Ikaros gene is required for the development of all lymphoid lineages. Cell. 1994; 79:143–56. [PubMed: 7923373]
14. Klug CA, et al. Hematopoietic stem cells and lymphoid progenitors express different Ikaros isoforms, and Ikaros is localized to heterochromatin in immature lymphocytes. Proc Natl Acad Sci U S A. 1998; 95:657–62. [PubMed: 9435248]
15. Mullighan CG, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med. 2009; 360:470–80. [PubMed: 19129520]
16. Mullighan CG, et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature. 2008; 453:110–4. [PubMed: 18408710]
17. Wilsker D, Patsialou A, Dallas PB, Moran E. ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. Cell Growth Differ. 2002; 13:95–106. [PubMed: 11959810]
18. Lahoud MH, et al. Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs. Genome Res. 2001; 11:1327–34. [PubMed: 11483573]

19. Chang LW, et al. Computational identification of the normal and perturbed genetic networks involved in myeloid differentiation and acute promyelocytic leukemia. Genome Biol. 2008; 9:R38. [PubMed: 18291030]

20. Akasaka T, et al. Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL). Blood. 2007; 109:3451–61. [PubMed: 17170124]

Europe PMC Funders Author Manuscripts

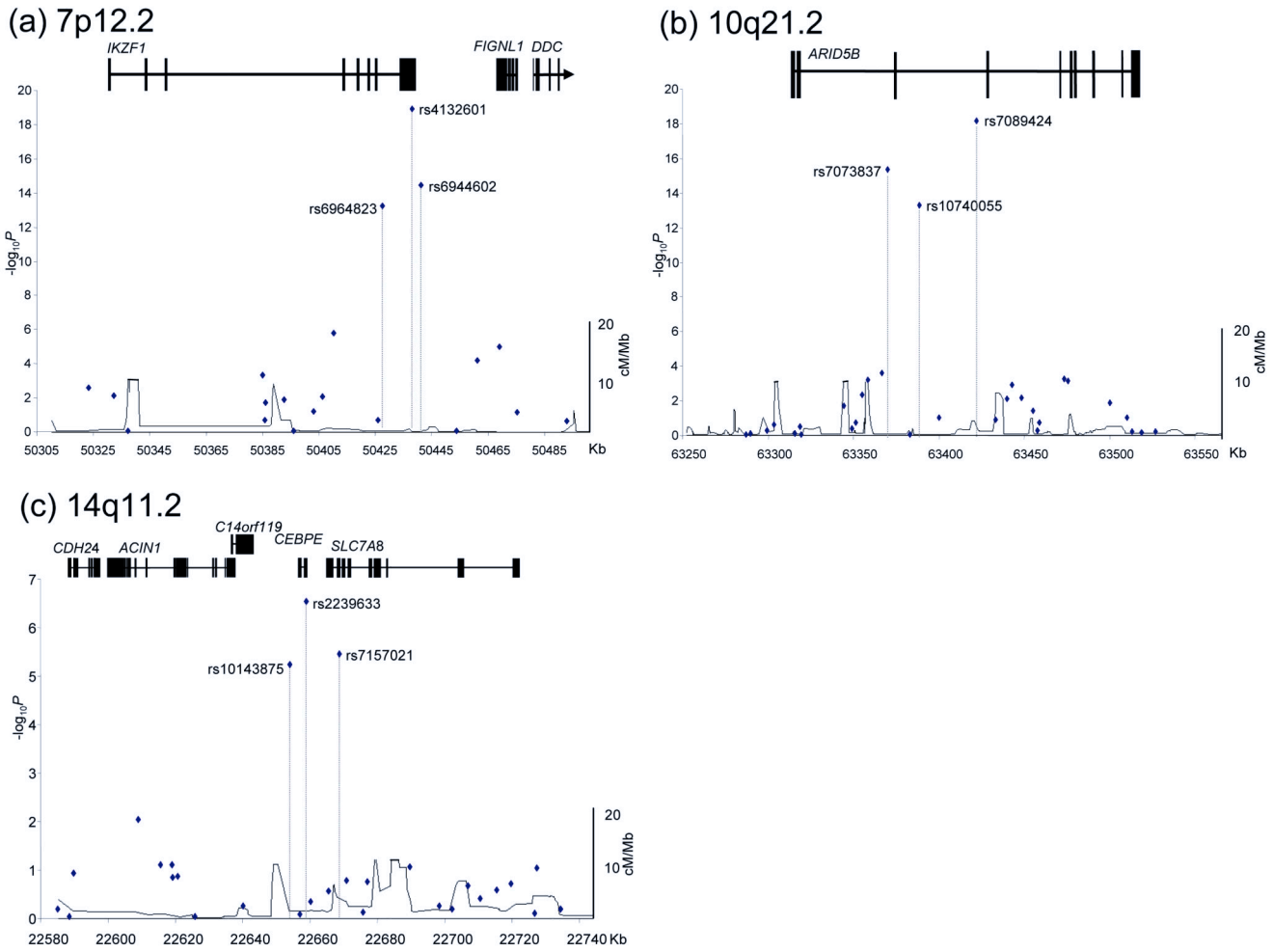

Figure 1. LD structure and association results for each of the disease-associated regions:

(A) 7p12.2; (B) 10q21.2 and (C) 14q11.2. Chromosomal positions based on NCBI build 36 coordinates, showing Ensemble (release 48) genes. Armitage trend test *P* values (as $-\log_{10}$ values; left y axis) are shown for SNPs analyzed. Recombination rates in HapMap CEU across the region are shown in black (right y axis). Also shown are the relative positions of genes mapping to each region of association. Exons of genes have been redrawn to show the relative positions in the gene, therefore maps are not to physical scale.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

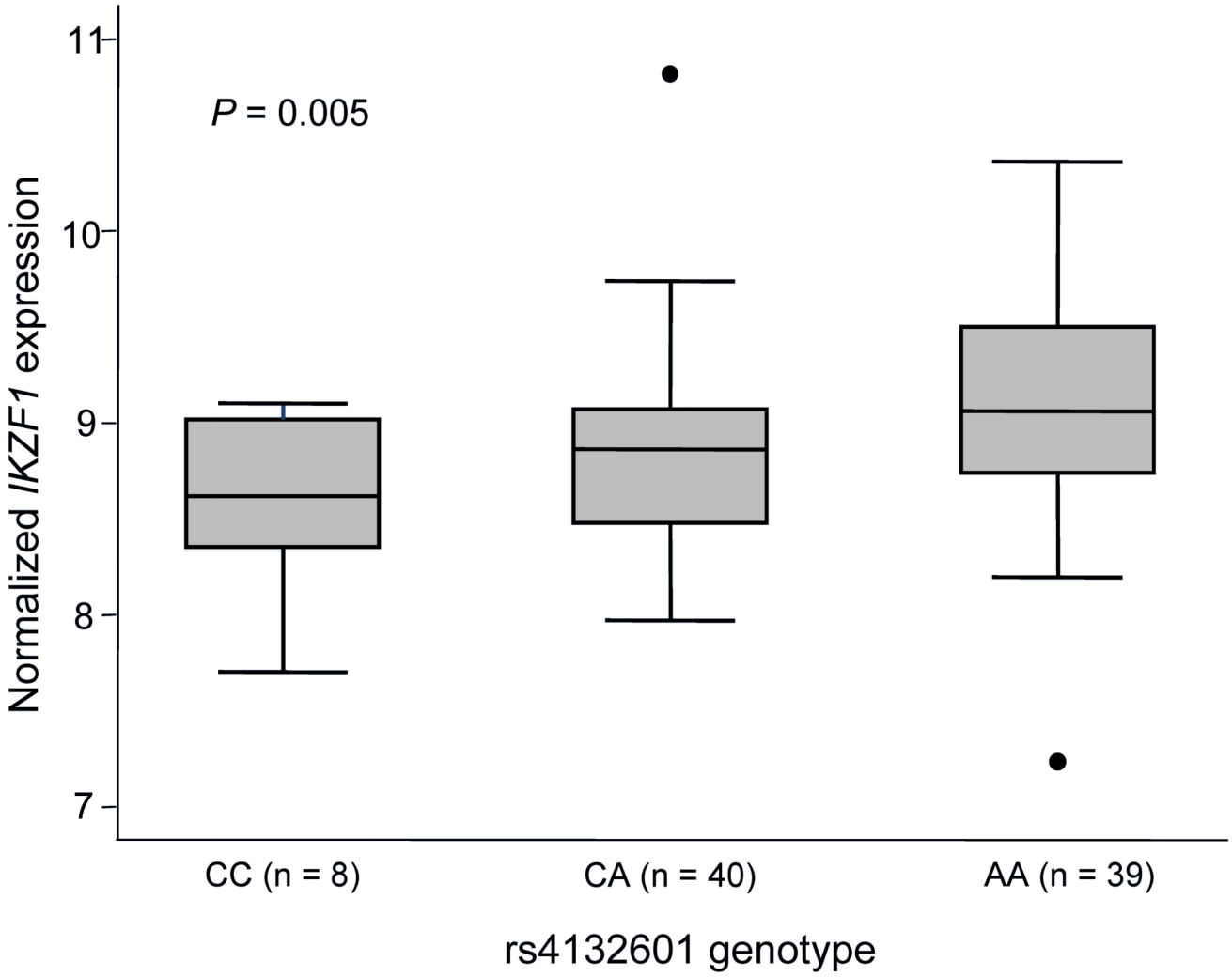

Figure 2. Relationship between lymphocyte mRNA expression levels of *IKZF1* and rs4132601 genotype.

Europe PMC Funders Author Manuscripts

Table 1

SNPs that meet a point-wise significance of P < 5 x 10^{-7}.

Results from analysis confined to B-cell ALL shown in parentheses.

SNP	Chr	Gene	Location (bps)	Risk allele[c]	Risk allele frequency	GWA-1	GWA-2	Combined			
						OR[a], 95% CI	OR[a], 95% CI	OR[a], 95% CI	*P* value[b]	$d_{P_{het}}$	e_{I^2}
rs6964823	7p12.2	*IKZF1*	50,427,590	G	0.5	1.49, 1.35 - 1.64 (1.48, 1.33 - 1.63)	1.57, 1.40 - 1.74 (1.61, 1.44 - 1.79)	1.52,1.41 - 1.64 (1.53, 1.42 - 1.65)	$6.02x10^{-14}$ ($1.88 x10^{-13}$)	0.671	0
rs4132601	7p12.2	*IKZF1*	50,438,098	C	0.28	1.66,1.51 - 1.81 (1.68, 1.52 - 1.83)	1.75, 1.57 - 1.93 (1.81, 1.63 - 2.00)	1.69, 1.58 - 1.81 (1.73, 1.61 - 1.85)	$1.20x10^{-19}$ ($9.31x10^{-20}$)	0.677	0
rs6944602	7p12.2	*IKZF1*	50,441,245	A	0.79	1.60, 1.44 - 1.76 (1.60, 1.43 - 1.76)	1.72, 1.52 - 1.91 (1.83, 1.63 - 2.04)	1.64, 1.37 - 2.07 (1.69, 1.56 - 1.81)	$3.42x10^{-15}$ ($1.51x10^{-15}$)	0.595	0
rs3779084	7p12.2	*DCC*	50,536,229	C	0.22	1.36, 1.20 - 1.53 (1.42, 1.25 - 1.59)	1.54, 1.35 - 1.73 (1.60, 1.41 - 1.80)	1.44, 1.32 - 1.56 (1.50, 1.37 - 1.63)	$8.81x10^{-9}$ ($6.50x10^{-10}$)	0.336	0
rs880028	7p12.2	*DCC*	50,537,630	**G**	0.22	1.35, 1.18 - 1.51 (1.40, 1.23 - 1.57)	1.54, 1.35 - 1.73 (1.60, 1.41 - 1.80)	1.43, 1.30 - 1.56 (1.49, 1.36 - 1.61)	$1.26x10^{-7}$ ($1.41x10^{-9}$)	0.286	12%
rs7809758	7p12.2	*DCC*	50,540,827	**G**	0.37	1.41, 1.26 - 1.56 (1.45, 1.30 - 1.60)	1.47, 1.30 - 1.64 (1.52, 1.34 - 1.70)	1.44, 1.32 - 1.54 (1.48, 1.37 - 1.60)	$2.41x10^{-10}$ ($2.88x10^{-11}$)	0.73	0
rs7073837	10q21.2	*ARIDB5*	63,369,901	A	0.4	1.61, 1.46 - 1.75 (1.62, 1.47 - 1.77)	1.54, 1.37 - 1.70 (1.55, 1.38 - 1.73)	1.58, 1.35 - 1.89 (1.59, 1.48 - 1.71)	$4.66x10^{-16}$ ($1.03x10^{-15}$)	0.692	0
rs10740055	10q21.2	*ARIDB5*	63,388,485	**C**	0.5	1.60, 1.45 - 1.74 (1.64, 1.48 - 1.79)	1.44, 1.27 - 1.61 (1.49, 1.31 - 1.66)	1.53, 1.41 - 1.64 (1.57, 1.45 - 1.81)	$5.35x10^{-14}$ ($1.61x10^{-14}$)	0.365	0
rs7089424	10q21.2	*ARIDB5*	63,422,165	C	0.34	1.74, 1.59 - 1.89 (1.78, 1.63 - 1.93)	1.54, 1.38 - 1.71 (1.56, 1.42 - 1.77)	1.65, 1.54 - 1.76 (1.70, 1.58 - 1.81)	$6.69x10^{-19}$ ($1.41x10^{-19}$)	0.292	7%
rs2239633	14q11.2	*CEBPE*	22,658,897	**G**	0.52	1.42, 1.27 - 1.57 (1.46, 1.30 - 1.61)	1.23, 1.07 - 1.40 (1.27, 1.10 - 1.45)	1.34, 1.22 - 1.45 (1.37, 1.26 - 1.49)	$2.88x10^{-7}$ ($5.60x10^{-8}$)	0.221	33%

[a] Odds ratio (OR) and 95% confidence interval per copy of risk allele.

[b] *P*-values denote Cochran-Armitage trend test statistics.

[c] Ancestral allele annotated by dbSNP embolded.

[d] P_{het} derived from Cochran's test of between study heterogeneity.

[e] I^2 denotes the proportion of the total variation due to heterogeneity; values ≥ 75% are considered characteristic of large heterogeneity.

Europe PMC Funders Author Manuscripts

Table 2

Relationship between 7p12.2-*IKZF1* (rs4132601), 10q21.2-*ARIDB5* (rs7089424) and 14q11.2-*CEBPE* (rs2239633) variants and ALL subtypes.

SNP	Gene	Chr	Risk allele	Risk allele frequency in controls	B, T lineage ALL			B lineage subtypes			
					B	T	*P* value	Hyperdiploid	TEL/AML1	Other	*P* value
rs4132601	*IKZF1*	7p12.2	C	0.27	0.4	0.33	0.076	0.41	0.38	0.41	0.711
rs7089424	*ARIDB5*	10q22.1	C	0.34	0.47	0.39	0.055	0.55	0.42	0.42	3.84×10^{-6}
rs2239633	*CEBPE*	14q11.2	G	0.52	0.6	0.51	0.016	0.61	0.61	0.6	0.783