Published in final edited form as:
Nat Genet. 2009 September ; 41(9): 1001–1005. doi:10.1038/ng.432.

Germline genomic variations associated with childhood acute lymphoblastic leukemia

Lisa R. Treviño[1,†], **Wenjian Yang**[1,†], **Deborah French**[1], **Stephen Hunger**[2], **William L. Carroll**[4], **Meenakshi Devidas**[3], **Cheryl Willman**[5], **Geoffrey Neale**[1], **James Downing**[1], **Susana Raimondi**[1], **Ching-Hon Pui**[1], **William E. Evans**[1], and **Mary V. Relling**[1]

[1]St. Jude Children's Research Hospital, Memphis TN, USA

[2]University of Colorado, Denver, CO USA

[3]University of Florida, Gainesville FL USA

[4]New York Medical Center, New York NY USA

[5]University of New Mexico, Albuquerque NM, USA

Abstract

We identified germline single nucleotide polymorphisms (SNPs) associated with childhood acute lymphoblastic leukemia (ALL) and its subtypes. Using the Affymetrix 500K Mapping array and publicly available genotypes, we identified 18 SNPs whose allele frequency differed ($P<1\times10^{-5}$) between a pediatric ALL population (n=317) and non-ALL controls (n=17,958). Six of these SNPs differed ($P\leq0.05$) in allele frequency among four ALL subtypes. Two SNPs in *ARID5B* not only differed between ALL and non-ALL groups (rs10821936, $P=1.4\times10^{-15}$, odds ratio[OR]=1.91; rs10994982, $P=5.7\times10^{-9}$, OR=1.62) but also distinguished B-hyperdiploid ALL from other subtypes (rs10821936, $P=1.62\times10^{-5}$, OR=2.17; rs10994982, P=0.003, OR 1.72). These *ARID5B* SNPs also distinguished B-hyperdiploid ALL from other subtypes in an independent validation cohort (n=124 children with ALL) (P=0.003 and P=0.0008, OR 2.45 and 2.86, respectively) and were associated with methotrexate accumulation and gene expression pattern in leukemic lymphoblasts. We conclude that germline genomic variations affect susceptibility to and characteristics of specific ALL subtypes.

Pediatric acute lymphoblastic leukemia (ALL) comprises biologically and clinically diverse subtypes. Somatically acquired genetic aberrations in ALL lymphoblasts are prognostic and can guide risk-directed therapy.[1] However, the extent to which germline variation contributes to susceptibility to ALL, to the acquisition of genetic aberrations that define ALL subtypes, and perhaps to the response to drug therapy among subtypes, is unknown. Candidate gene approaches have identified inherited polymorphisms in loci that may contribute to susceptibility to ALL, including the multidrug resistance gene *ABCB1/MDR1*, methylenetetrahydrofolate reductase (*MTHFR*), the glutathione-S-transferases as well as cell-cycle inhibitor and DNA mismatch repair genes.[2-7] Lacking, however, are genome-wide studies that assess how inherited variation contributes to the development of ALL. Therefore, we conducted a genome-wide association study to identify germline single nucleotide

Address correspondence and reprint requests to Mary V. Relling, PharmD, St. Jude Children's Research Hospital, 332 N Lauderdale, Memphis, TN 38105-2794, USA. Phone: 901-495-2348; Fax: 901-525-6869; mary.relling@stjude.org.
†Drs Treviño and Yang contributed equally to this body of work.

Author contributions L.T., W.Y. and M.V.R designed the study. M.V.R. and W.E.E. directed the study. L.T., W.Y.,D.F., S.H., W.L.C., M.D., C.W., G.N., J.D., S.R., C.H.P, W.E.E., M.V.R. helped with data analysis and data interpretation. W.Y. conducted the statistical genomic analysis. S.R. performed and guided cytogenetic data analysis. S.H., W.L.C, M.D., and C.H.P. provided clinical data. J.D. provided leukemic lymphoblast gene expression data.

polymorphisms (SNPs) in children with newly diagnosed ALL that may be associated with the development of ALL and with specific ALL subtypes.

We first identified SNPs whose allele frequency differed between a discovery cohort of 317 children of European descent with ALL and 17,958 individuals of European descent without ALL from three independent control groups (Supplementary Fig. 1). After applying quality control criteria, we evaluated 307,944 germline SNPs. Eighteen SNPs differed in allele frequency (using P-value thresholds specified in Supplementary Fig. 1) between patients with ALL and non-ALL controls (Table 1, Fig. 1); they were annotated to 12 unique genes, with odds ratios ranging from 1.43 to 3.62. Two of the 18 SNPs were annotated to the AT-rich interactive domain 5B *(ARID5B)* gene (rs10821936, $P=1.4\times10^{-15}$ and rs10994982, $P=5.7\times10^{-9}$) and were in linkage disequilibrium (LD) with one another ($r^2 = 0.42$, $P<1\times10^{-10}$). Three SNPs were annotated to chromosomal region 7p12.2, including one annotated to the zinc finger protein subfamily 1A (*IKZF1*) gene (rs11978267, $P=8.8\times10^{-11}$) and two annotated to the dopa decarboxylase aromatic L-amino acid (*DDC*) gene (rs2167364, $P=2.8\times10^{-6}$ and rs2242041, $P=9.9\times10^{-7}$). These three SNPs were in LD with each other (pairwise $r^2>0.28$, $P<1\times10^{-10}$).

We next compared the allele frequency of these 18 SNPs among four major ALL subtypes (B-other, B-hyperdiploid, t(12;21)/*ETV6-RUNX1*, and T-cell ALL) in the discovery cohort and found that six SNPs distinguished among the subtypes ($P\leq0.05$) (Table 2). The two *ARID5B* SNPs (rs10821936 and rs10994982) distinguished B-hyperdiploid ALL from all other subtypes; for example, the frequency of the C allele at rs10821936 was greater in patients with B-hyperdiploid ALL (61%) than in all other patients with ALL (42%; $P=1.62\times10^{-5}$) and than in non-ALL controls (33%) (Table 2, Table 3). The three SNPs localized to 7p12.2 distinguished T-cell ALL from the three other subtypes ($P\leq0.020$). One SNP annotated to *OR2C3* was associated with the t(12;21)/*ETV6-RUNX1* subtype (P=0.021).

Two of the six SNPs that distinguished ALL subtypes in the discovery cohort also distinguished subtypes in a validation cohort of children of European descent with ALL (n=124). Both were annotated to *ARID5B*, both distinguished B-hyperdiploid ALL from other subtypes (Table 2, Table 3) and both were confirmed with an alternative genotyping methodology with greater than 99% accuracy (details in Methods). A haplotype analysis in the combined patient cohort (discovery plus validation cohorts) revealed that haplotypes including both *ARID5B* SNPs distinguished patients with B-hyperdiploid ALL from other ALL patients (score test $P=1.0\times10^{-6}$; Supplementary Table 1). The allele frequencies of the remaining four SNPs (Table 2) were not significantly different among the ALL subtypes in the validation cohort after correction for multiple testing, although some SNPs (e.g., rs2167364, P=0.03) displayed a trend toward subtype association.

The *ARID5B* SNPs were associated with B-hyperdiploid ALL, which has a better response to methotrexate chemotherapy than other ALL subtypes. [8] Because this response is partly due to greater accumulation of methotrexate polyglutamates in B-hyperdiploid than in non-B-hyperdiploid ALL blast cells,[9] we investigated whether these SNPs were associated with the clinical phenotype of methotrexate polyglutamate accumulation. We found the same alleles of *ARID5B* SNPs (rs10821936 and rs10994982) that were associated with the likelihood of B-hyperdiploid ALL were also associated with greater methotrexate polyglutamate accumulation (P=0.005 and P=0.021, respectively) in patients with this ALL subtype (n=37; Supplementary Fig. 2A and 2B). Both SNPs were also associated with the expression phenotype of global gene expression pattern in B-hyperdiploid ALL blast cells (n=44, Supplementary Fig. 3). The expression of eight genes was associated with *ARID5B* SNP genotype ($P<5.3\times10^{-5}$, false discovery rate ≤ 10%, Supplementary Table 2). Of these, the *MKL1* gene, which encodes the

megakaryoblastic leukemia 1 protein, was most strongly associated with the *ARID5B* SNP genotypes ($P = 1.84 \times 10^{-6}$). Similar associations were not observed in the other ALL subtypes.

Cancer subtype-specific genomic variation has been shown to be important in breast,[10] lung, [11], gastric[12] and myeolproliferative[13] cancers. Our study represents the first genome-wide interrogation to identify genetic risk factors for ALL, the most common childhood cancer. Because childhood ALL is a biologically heterogeneous disease with molecular subtypes that differ in response to chemotherapy and prognosis,[14] we used a genome-wide analysis to identify risk factors not only for ALL but also for the main ALL subtypes.

Epidemiologic studies have identified putative environmental[14] and genetic[15] risk factors for ALL, although most display only a modest association with disease risk, and few studies have examined specific ALL subtypes. Only a small percentages of ALL cases are associated with Mendelian diseases and genetic syndromes such as Down syndrome, ataxia-telangiectasia, and β-thalassemia.[16-18] The greater risk of ALL in children than in adults has been linked to developmental immaturity of the immune system[19] and differential exposure to environmental toxins.[20] Because children have less cumulative exposure to mutagens than adults, some hypothesize that genetic predisposition to cancer may be greater in children than in adults. Candidate gene approaches have identified polymorphisms in the carcinogen metabolism genes *GSTM1*, *GSTT1*, and *CYP1A1*[21] as well as the cell cycle checkpoint genes *CDKN1B*, *CDKN2A* and 2B[6] (Supplementary Table 3) that may predispose to ALL. Polymorphisms in the *MTHFR* and *NQO1* genes have also been associated with distinct subtypes of childhood leukemia (B-hyperdiploid ALL and leukemias with *MLL* rearrangements, respectively).[4] Several of these candidate genes were interrogated in our analysis. Although we acknowledge the limitations of our genotyping platforms, as well as differences in study cohorts, we did not find that variants in previously reported genes were associated with the risk of ALL in our study (Supplementary Table 3).

We used a two-step genome-wide approach to reveal novel associations between gene polymorphisms and the risk of specific subtypes of ALL. We found 18 germline SNPs whose allele frequency differed between children with ALL, all of European descent, and three independent European and American Caucasian control groups (Table 1). We confirmed that comparing SNP data from externally-genotyped controls to children with ALL typed at our center was unlikely to be biased by underlying population stratification [quantile-quantile (Q-Q) plot, genomic control lambda parameter = 1; Supplementary Fig. 4]. Of these 18 SNPs, we tested for differences in allele frequency among the major ALL subtypes. By using the initial case-control study as a screen, we were able to narrow our attention on polymorphisms that not only distinguished subtypes but also might contribute to the overall risk of ALL. This two-step approach also allowed us to capture additional SNPs that may have been overlooked had we limited our analysis to subtype-related differences in SNP genotypes solely within the 317 children with ALL (Supplementary Tables 4-7).

Two *ARID5B* SNPs (rs10821936 and rs10994982, Table 2) discriminated B-hyperdiploid ALL from non-ALL controls and other ALL subtypes. These SNPs also showed significant association with two phenotypes in B-hyperdiploid ALL: the clinical phenotype of intracellular accumulation of methotrexate polyglutamates (Supplementary Fig. 2A and 2B) and the expression phenotype of global gene expression pattern in ALL blast cells (Supplementary Fig. 3).

Both SNPs were located within intron 3 of the *ARID5B* gene and were encompassed by an LD block that spanned exons 3 and 4 (Supplementary Fig. 5). To determine whether they might exert a functional effect in *ARID5B* via LD with coding polymorphisms in *ARID5B*, we sequenced exon 3 and exon 4 in 63 HapMap CEPH cell lines. However, we did not identify

any coding SNPs in these regions. We acknowledge that further functional analyses are required to elucidate the mechanism by which these two SNPs may affect risk of ALL.

The *ARID5B* gene (also known as *DESRT* and *MRF2*) is a member of the ARID family of transcription factors and plays important roles in embryonic development, cell-type—specific gene expression, and cell growth regulation.[22] Homozygous knockout mice (*desrt* $^{-/-}$) display abnormal thymic and splenic architecture and disrupted B cell differentiation.[23-25] *ARID5B* expression is upregulated in patients with acute megakaryoblastic leukemia[26] and acute promyelocytic leukemia.[27] Thus, it is possible that germline variation at the *ARID5B* locus affects susceptibility to this B-lineage leukemia by altering *ARID5B* function in B-lineage development.

Three of the top 18 SNPs, localized to the genes *IKZF1* and *DDC* in the 7p12.2 chromosomal region, distinguished T-cell from B-lineage ALL. *IKZF1* encodes Ikaros, critical for normal mouse and human lymphoid development[28, 29] and whose deletion contributes to the pathogenesis of a very aggressive form of childhood ALL.[30-33] However, the association of germline *IKZF1* and *DDC* SNPs with T-cell ALL was not replicated in our validation cohort, possibly due to the smaller number of patients.

Our findings indicate that inherited genetic variation contributes to the risk of childhood ALL and is likely to contribute to the development of specific ALL subtypes. The data further suggest that the same genetic variation that predisposes to B-hyperdiploid ALL may underlie the superior response of this subtype to chemotherapy. Thus, genomic variation may affect not only disease risk but treatment outcome as well.

Methods

Patients and DNA samples

We analyzed germline DNA from 441 children of European descent with the four most common ALL subtypes (see below). The discovery cohort consisted of 317 patients of European descent, including 262 patients from the St. Jude Children's Research Hospital Total XIIIB (1994-1998) and XV (2000-2007) ALL protocols[34, 35] and 55 patients (all with B-precursor ALL subtypes) from the Children's Oncology Group 9906 study.[36] The validation cohort consisted of the next 124 children of European descent enrolled on the St. Jude Total XV ALL protocol. This study was approved by the St. Jude Institutional Review Board and signed informed consent was obtained from patients, parents, or guardians, as appropriate.

External data sets

We obtained SNP allele frequency and genotype data from three groups of European descent to serve as non-ALL controls. One group comprised participants in the Wellcome Trust Case Control Consortium (WTCCC; n=14,311; http://www.wtccc.uk.org),[37] excluding those with bipolar disease. The other two groups were the Genetic Association Informative Network (GAIN; http://www.genome.gov/19518664) schizophrenia (phs000021.v1.p1; n=2,601) and bipolar disorder (phs000017.v1.p1; n=1,046)[38-40] study cohorts. Because the prevalence of adult survivors of childhood ALL is less than 1:1,000, these three groups were considered non-ALL controls.

Characterization of ALL molecular and immunophenotypic subgroups

The immunophenotyping and genotyping of leukemic lymphoblasts from St. Jude patients[34] and COG patients were previously described.[36, 41] We analyzed only patients with the four most common, non-overlapping ALL subtypes: B-lineage ALL with no defined genetic or chromosomal abnormalities (B-other; n=121 in discovery cohort, n=44 in validation cohort);

B-lineage hyperdiploid ALL with more than 50 chromosomes (B-hyperdiploid; n=108 and n=36); B-lineage ALL bearing the t(12;21)/*ETV6-RUNX1* fusion (n=45 and n=20); and T-cell ALL (n=43 and n=24) (Supplementary Table 8 online).

Genotyping and SNP Filtering Criteria

DNA was extracted from the blood of patients with ALL during complete remission. For patients in the discovery cohort and for 65 patients in the validation cohort, 500 ng of DNA was digested with *Nsp* and *Sty* restriction enzymes for the 500K Array Set chips (Affymetrix, Santa Clara, CA). DNA was amplified, labeled, and hybridized to chips as described.[30] The chips were scanned, and genotype calls were made by using the Bayesian Robust Linear Multichip with Mahalanobis Distance (BRLMM) algorithm for a total of 500,568 possible SNPs interrogated. Genotyping for 59 patients in the validation cohort was performed by using the Affymetrix Genome-Wide Human SNP Array 6.0 based on the Birdseed genotype-calling algorithm, which overlaps with 482,251 of the 500,568 SNPs on the 500K chip. Visual inspection of original allele-specific signal intensity plots (i.e. theta plots) was carried out to ensure that genotyping calls clustered distinctly. Genotyping for participants in the Wellcome Trust Case Control Consortium study was performed as reported, using the Affymetrix 500K Mapping Array sets.[37] The GAIN participants were analyzed by using the Affymetrix Genome-Wide Human SNP Array 6.0. The confirmation of *ARID5B* SNP genotypes in 386 ALL cases was performed using iPLEX from Sequenom, Inc., in the University of Chicago's Genetic Services Laboratories.

SNPs with a genotyping call rate < 96% (n=84,032) in both the St. Jude and COG data sets were excluded in the discovery cohort. Of the remaining 416,536 SNPs interrogated, only those with adequate quality control measures in the WTCCC cohort [37] and both GAIN cohorts (http://www.genome.gov/19518664) were included in the genome-wide association analysis (n=307,944; Supplementary Fig. 1).

Sequencing of exon 3 and exon 4

Ten nanograms of genomic DNA from 63 of the CEPH HapMap samples was used to amplify and sequence exon 3 and exon 4 of the *ARID5B* gene (Supplementary Table 9). Sequencing of the amplified product was performed using the dye-terminator chemistry (Applied Biosystems) in the St. Jude Children's Research Hospitals' Hartwell Center for Bioinformatics.

Measurement of methotrexate polyglutamate accumulation

In vivo intracellular accumulation of methotrexate polyglutamate metabolites was measured as previously described[42-44] in bone marrow leukemic lymphoblasts from 118 patients. Bone marrow samples were obtained 42-44 hours after a single treatment with high-dose methotrexate (1 g/m^2 given intravenously over 4 or 24 hours) during the "upfront window" before remission induction therapy.[34, 35]

Gene expression profiling

Total RNA was extracted (TriReagent, MRC, Cincinnati, OH, USA) from cryopreserved mononuclear cell suspensions from bone marrow obtained from 156 patients of European descent at the time of diagnosis of ALL. The Affymetrix HG-U133A 2.0 Array (Affymetrix Inc, Santa Clara, CA, USA), comprising more than 22,283 probe sets, was used to interrogate the expression of RNA as described.[45, 46] Gene expression data were analyzed by using the Affymetrix MAS5.0 algorithm.

Statistical analysis

The ancestry of patients with ALL in both the discovery and validation cohorts was inferred on the basis of the approximately 200,000 SNPs with a call rate > 99%. We used the genotypes of samples from the International HapMap project (www.hapmap.org; Phase II; 210 unrelated individuals of known ancestry) as a reference population and used STRUCTURE[47] to assess the percentage of European, African, and Asian ancestry for each patient. Patients whose ancestry was greater than 90% European were included in the study; however, almost identical SNP associations were observed when we used more stringent criteria (95% vs. 90%) to define European ancestry (Supplementary Table 10).

R 2.6.1 statistical software (http://www.r-project.org/) was used for analysis. For each SNP, logistic regression was used to compare the frequency of the B allele (e.g., for genotypes AA, AB, and BB, the frequency of the B allele is 0, 1, and 2) between the ALL patient population and the combined control group and between the ALL patient population and each of the three non-ALL control groups. A multinomial log-linear model was used to compare the allele frequencies of selected SNPs among the four ALL subtypes. Logistic regression was also used to compare allele frequencies between single ALL subtypes and all other subtypes combined (e.g., B-hyperdiploid ALL vs. non-B-hyperdiploid ALL; Supplementary Fig. 1). All odds ratios reported are allelic odds ratios unless noted otherwise. All P values are reported as two-tailed P values.

The association between SNPs in *ARID5B* and leukemia-cell gene expression was analyzed by multiple linear regression for each of the 22,283 probe sets on the Affymetrix HG-U133A 2.0 Array (Affymetrix Inc, Santa Clara, CA, USA). The probe sets were rank-ordered by their significance in the regression model as assessed by analysis of variance (ANOVA). False discovery rates were estimated by using the q-value method. Multiple linear regression analysis was also used to compare SNP genotypes with methotrexate polyglutamate accumulation.

Data analysis overview

In the first step, we compared the allele frequencies of 307,944 SNP genotypes in the discovery cohort of 317 patients with ALL and the non-ALL control groups (Supplementary Fig. 1). For the second step, SNPs whose allele frequency differed between patients in the ALL and combined non-ALL group ($P < 1.0 \times 10^{-5}$) and between the ALL group and each of the three individual non-ALL groups ($P < 0.01$) were then compared among the four ALL subtypes. SNPs that distinguished among ALL subtypes were then tested in an independent validation cohort of 124 children of European descent with ALL. As a secondary analysis, we compared SNP genotypes only within ALL cases in the discovery cohort that distinguished among the four main ALL subtypes, without considering the non-ALL controls (Supplementary Tables 4-7).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank our protocol co-investigators, clinical and research staff (particularly Ms. Pamela McGill, Dr. Jean Cai, Dr. Shih-Hsiang Chen, Ms. Nancy Kornegay and Dr. Jennifer Pauley), the Hartwell Center for Bioinformatics and Biotechnology, Dr. Soma Das at the Genetic Services Laboratory at the University of Chicago and the patients and their families who participated. We thank Dr. Robert Whitson from City of Hope Hospital for constructive discussions.

This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113.

Funding support for the Genome-Wide Association of Schizophrenia Study was provided by the National Institute of Mental Health (R01 MH67257, R01 MH59588, R01 MH59571, R01 MH59565, R01 MH59587, R01 MH60870, R01 MH59566, R01 MH59586, R01 MH61675, R01 MH60879, R01 MH81800, U01 MH46276, U01 MH46289 U01 MH46318, U01 MH79469, and U01 MH79470) and the genotyping of samples was provided through the Genetic Association Information Network (GAIN). The datasets used for the analyses described in this manuscript were obtained from the database of Genotype and Phenotype database (dbGaP) found at http://www.ncbi.nlm.nih.gov/gap through dbGaP (accession number phs000021.v1.p1). Samples and associated phenotype data for the Genome-Wide Association of Schizophrenia Study were provided by the Molecular Genetics of Schizophrenia Collaboration (PI: Pablo V. Gejman, Evanston Northwestern Healthcare (ENH) and Northwestern University, Evanston, IL, USA).

Funding support for the Whole Genome Association Study of Bipolar Disorder was provided by the National Institute of Mental Health (NIMH) and the genotyping of samples was provided through the Genetic Association Information Network (GAIN). The datasets used for the analyses described in this manuscript were obtained from the database of Genotype and Phenotype database (dbGaP) found at http://www.ncbi.nlm.nih.gov/gap through dbGaP (accession number phs000017.v1.p1). Samples and associated phenotype data for the Collaborative Genomic Study of Bipolar Disorder were provided by The NIMH Genetics Initiative for Bipolar Disorder. Data and biomaterials were collected in four projects that participated in the NIMH Bipolar Disorder Genetics Initiative. From 1991-98, the Principal Investigators and Co-Investigators were: Indiana University, Indianapolis, IN, U01 MH46282, John Nurnberger, M.D., Ph.D., Marvin Miller, M.D., and Elizabeth Bowman, M.D.; Washington University, St. Louis, MO, U01 MH46280, Theodore Reich, M.D., Allison Goate, Ph.D., and John Rice, Ph.D.; Johns Hopkins University, Baltimore, MD U01 MH46274, J. Raymond DePaulo, Jr., M.D., Sylvia Simpson, M.D., MPH, and Colin Stine, Ph.D.; NIMH Intramural Research Program, Clinical Neurogenetics Branch, Bethesda, MD, Elliot Gershon, M.D., Diane Kazuba, B.A., and Elizabeth Maxwell, M.S.W. Data and biomaterials were collected as part of ten projects that participated in the NIMH Bipolar Disorder Genetics Initiative. From 1999-03, the Principal Investigators and Co-Investigators were: Indiana University, Indianapolis, IN, R01 MH59545, John Nurnberger, M.D., Ph.D., Marvin J. Miller, M.D., Elizabeth S. Bowman, M.D., N. Leela Rau, M.D., P. Ryan Moe, M.D., Nalini Samavedy, M.D., Rif El-Mallakh, M.D. (at University of Louisville), Husseini Manji, M.D. (at Wayne State University), Debra A. Glitz, M.D. (at Wayne State University), Eric T. Meyer, M.S., Carrie Smiley, R.N., Tatiana Foroud, Ph.D., Leah Flury, M.S., Danielle M. Dick, Ph.D., Howard Edenberg, Ph.D.; Washington University, St. Louis, MO, R01 MH059534, John Rice, Ph.D, Theodore Reich, M.D., Allison Goate, Ph.D., Laura Bierut, M.D. ; Johns Hopkins University, Baltimore, MD, R01 MH59533, Melvin McInnis M.D. , J. Raymond DePaulo, Jr., M.D., Dean F. MacKinnon, M.D., Francis M. Mondimore, M.D., James B. Potash, M.D., Peter P. Zandi, Ph.D, Dimitrios Avramopoulos, and Jennifer Payne; University of Pennsylvania, PA, R01 MH59553, Wade Berrettini M.D.,Ph.D. ; University of California at Irvine, CA, R01 MH60068, William Byerley M.D., and Mark Vawter M.D. ; University of Iowa, IA, R01 MH059548, William Coryell M.D. , and Raymond Crowe M.D. ; University of Chicago, IL, R01 MH59535, Elliot Gershon, M.D., Judith Badner Ph.D. , Francis McMahon M.D. , Chunyu Liu Ph.D., Alan Sanders M.D., Maria Caserta, Steven Dinwiddie M.D., Tu Nguyen, Donna Harakal; University of California at San Diego, CA, R01 MH59567, John Kelsoe, M.D., Rebecca McKinney, B.A.; Rush University, IL, R01 MH059556, William Scheftner M.D. , Howard M. Kravitz, D.O., M.P.H., Diana Marta, B.S., Annette Vaughn-Brown, MSN, RN, and Laurie Bederow, MA; NIMH Intramural Research Program, Bethesda, MD, 1Z01MH002810-01, Francis J. McMahon, M.D., Layla Kassem, PsyD, Sevilla Detera-Wadleigh, Ph.D, Lisa Austin,Ph.D, Dennis L. Murphy, M.D.

Supported by NCI grants CA 51001, CA 078224 , CA 36401 and CA 21765 and the NIH/NIGMS Pharmacogenetics Research Network and Database (U01 GM61393, U01 HL65899, U01GM61374 http://pharmgkb.org), by a Center of Excellence grant from the State of Tennessee, and by the American Lebanese Syrian Associated Charities (ALSAC).

References

1. Pui CH, Evans WE. Treatment of acute lymphoblastic leukemia. N. Engl. J Med 2006;354:166–178. [PubMed: 16407512]
2. Schnakenberg E, et al. Polymorphisms of methylenetetrahydrofolate reductase (MTHFR) and susceptibility to pediatric acute lymphoblastic leukemia in a German study population. BMC. Med. Genet 2005;6:23. [PubMed: 15921520]
3. Urayama KY, et al. MDR1 gene variants, indoor insecticide exposure, and the risk of childhood acute lymphoblastic leukemia. Cancer Epidemiol. Biomarkers Prev 2007;16:1172–1177. [PubMed: 17548681]
4. Wiemels JL, et al. Methylenetetrahydrofolate reductase (MTHFR) polymorphisms and risk of molecularly defined subtypes of childhood acute leukemia. Proceedings of the National Academy of Sciences of the United States of America 2001;98:4004–4009. [PubMed: 11274424]
5. Stanulla M, et al. Polymorphisms within glutathione S-transferase genes (GSTM1, GSTT1, GSTP1) and risk of relapse in childhood B-cell precursor acute lymphoblastic leukemia: a case-control study. Blood 2000;95:1222–1228. [PubMed: 10666194]

6. Healy J, et al. Promoter SNPs in G1/S checkpoint regulators and their impact on the susceptibility to childhood leukemia. Blood 2007;109:683–692. [PubMed: 17008550]
7. Mathonnet G, Krajinovic M, Labuda D, Sinnett D. Role of DNA mismatch repair genetic polymorphisms in the risk of childhood acute lymphoblastic leukaemia. Br. J. Haematol 2003;123:45–48. [PubMed: 14510941]
8. Paulsson K, et al. Evidence for a single-step mechanism in the origin of hyperdiploid childhood acute lymphoblastic leukemia. Genes Chromosomes. Cancer 2005;44:113–122. [PubMed: 15942938]
9. Kager L, et al. Folate pathway gene expression differs in subtypes of acute lymphoblastic leukemia and influences methotrexate pharmacodynamics 10. J. Clin. Invest 2005;115:110–117. [PubMed: 15630450]
10. Nordgard SH, et al. Genome-wide analysis identifies 16q deletion associated with survival, molecular subtypes, mRNA expression, and germline haplotypes in breast cancer patients. Genes Chromosomes. Cancer 2008;47:680–696. [PubMed: 18398821]
11. Sobti RC, et al. Combined effect of GSTM1, GSTT1 and GSTP1 polymorphisms on histological subtypes of lung cancer. Biomarkers 2008;13:282–295. [PubMed: 18415801]
12. Kim MS, et al. Frameshift mutation of UVRAG, an autophagy-related gene, in gastric carcinomas with microsatellite instability. Hum. Pathol 2008;39:1059–1063. [PubMed: 18495205]
13. Campbell PJ. Somatic and germline genetics at the JAK2 locus. Nat. Genet 2009;41:385–386. [PubMed: 19338077]
14. Greaves M. Childhood leukaemia. BMJ 2002;324:283–287. [PubMed: 11823363]
15. Chen HC, et al. Genetic polymorphisms of metabolic enzymes CYP1A1, CYP2D6, GSTM1 and GSTT1 and leukemia susceptibility. Eur. J. Cancer Prev 2008;17:251–258. [PubMed: 18414197]
16. Pui CH, Evans WE. Acute lymphoblastic leukemia. N. Engl. J. Med 1998;339:605–615. [PubMed: 9718381]
17. Pui CH, Relling MV, Downing JR. Acute lymphoblastic leukemia. N. Engl. J. Med 2004;350:1535–1548. [PubMed: 15071128]
18. Ziino O, et al. Acute lymphoblastic leukemia in children with associated genetic conditions other than Down's syndrome. The AIEOP experience. Haematologica 2006;91:139–140. [PubMed: 16434385]
19. Greaves MF. Speculations on the cause of childhood acute lymphoblastic leukemia. Leukemia 1988;2:120–125. [PubMed: 3278171]
20. Perera FP. Environment and cancer: who are susceptible? Science 1997;278:1068–1073. [PubMed: 9353182]
21. Dong LM, et al. Genetic susceptibility to cancer: the role of polymorphisms in candidate genes. JAMA 2008;299:2423–2436. [PubMed: 18505952]
22. Patsialou A, Wilsker D, Moran E. DNA-binding properties of ARID family proteins. Nucleic Acids Res 2005;33:66–80. [PubMed: 15640446]
23. 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;11:1327–1334. [PubMed: 11483573]
24. 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]
25. Wilsker D, et al. The DNA-binding properties of the ARID-containing subunits of yeast and mammalian SWI/SNF complexes. Nucleic Acids Res 2004;32:1345–1353. [PubMed: 14982958]
26. Bourquin JP, et al. Identification of distinct molecular phenotypes in acute megakaryoblastic leukemia by gene expression profiling. Proc. Natl. Acad. Sci. U. S. A 2006;103:3339–3344. [PubMed: 16492768]
27. 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]
28. Molnar A, Georgopoulos K. The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. Mol. Cell Biol 1994;14:8292–8303. [PubMed: 7969165]

29. Molnar A, et al. The Ikaros gene encodes a family of lymphocyte-restricted zinc finger DNA binding proteins, highly conserved in human and mouse. J. Immunol 1996;156:585–592. [PubMed: 8543809]
30. Mullighan CG, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature 2007;446:758–764. [PubMed: 17344859]
31. Mullighan CG, et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature 2008;453:110–114. [PubMed: 18408710]
32. den Boer ML, et al. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. Lancet Oncol. 2009
33. Mullighan CG, et al. Deletion of IKZF1 and Prognosis in Acute Lymphoblastic Leukemia. N. Engl. J. Med. 2009
34. Pui CH, et al. Improved outcome for children with acute lymphoblastic leukemia: results of Total Therapy Study XIIIB at St Jude Children's Research Hospital. Blood 2004;104:2690–2696. [PubMed: 15251979]
35. Pui C-H, et al. Treating childhood acute lymphoblastic leukemia without cranial irradiation. New England Journal of Medicine 2009;360:2730–2741. [PubMed: 19553647]
36. Borowitz MJ, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. Blood 2008;111:5477–5485. [PubMed: 18388178]
37. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678. [PubMed: 17554300]
38. Dick DM, et al. Genomewide linkage analyses of bipolar disorder: a new sample of 250 pedigrees from the National Institute of Mental Health Genetics Initiative. Am. J. Hum. Genet 2003;73:107–114. [PubMed: 12772088]
39. McInnis MG, et al. Genome-wide scan and conditional analysis in bipolar disorder: evidence for genomic interaction in the National Institute of Mental Health genetics initiative bipolar pedigrees. Biol. Psychiatry 2003;54:1265–1273. [PubMed: 14643094]
40. Suarez BK, et al. Genomewide linkage scan of 409 European-ancestry and African American families with schizophrenia: suggestive evidence of linkage at 8p23.3-p21.2 and 11p13.1-q14.1 in the combined sample. Am. J. Hum. Genet 2006;78:315–333. [PubMed: 16400611]
41. Borowitz MJ, et al. Minimal residual disease detection in childhood precursor-B-cell acute lymphoblastic leukemia: relation to other risk factors. A Children's Oncology Group study. Leukemia 2003;17:1566–1572. [PubMed: 12886244]
42. Fabre G, et al. In vitro formation of polyglutamyl derivatives of methotrexate and 7-hydroxymethotrexate in human lymphoblastic leukemia cells. Cancer Res 1983;43:4648–4652. [PubMed: 6192907]
43. Masson E, et al. Accumulation of methotrexate polyglutamates in lymphoblasts is a determinant of antileukemic effects in vivo. A rationale for high- dose methotrexate. J. Clin. Invest 1996;97:73–80. [PubMed: 8550853]
44. Synold TW, et al. Blast cell methotrexate-polyglutamate accumulation in vivo differs by lineage, ploidy, and methotrexate dose in acute lymphoblastic leukemia. J. Clin. Invest 1994;94:1996–2001. [PubMed: 7525652]
45. Ross ME, et al. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. Blood 2003;102:2951–2959. [PubMed: 12730115]
46. Yeoh EJ, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell 2002;1:133–143. [PubMed: 12086872]
47. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155:945–959. [PubMed: 10835412]

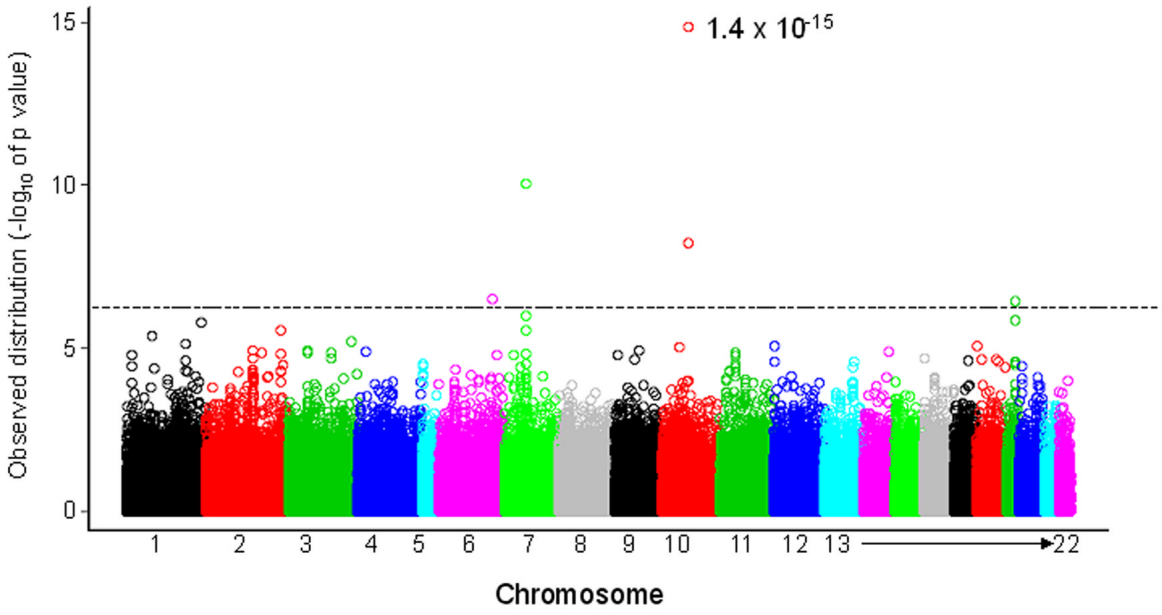

Figure 1. Genome-wide P values comparing allele frequency of SNPs between the ALL and combined non-ALL groups according to chromosome

The x axis shows the P values ($-\log_{10}$) for 307,944 germline SNPs. The allele frequency of the SNPs was compared between patients with ALL in the discovery cohort (n=317) and the combined non-ALL control cohort (n=17,958). The SNP whose allele frequency differed most significantly between the two cohorts ($P = 1.4 \times 10^{-15}$) was localized to chromosome 10. The dashed line shows the threshold P value indicating genome-wide significance.

Table 1

Germline SNPs whose allele frequencies differed between patients with ALL and non-ALL control groups.

dbSNP ID	Gene name	Chr	Alleles (A/B)[a]	Allelic OR (95% CI)[b]	Genomic OR (95% CI)[c]		P value[d]				Allele Frequency[e]			
					A/B-A/A	B/B-A/A	ALL vs. combined control group	ALL vs. WTCCC	ALL vs. GAIN: Schiz.	ALL vs. GAIN: bipolar	ALL (HWE P value)	WTCCC	Schiz	Bipolar
rs10821936	*ARID5B*	10	T/C*	1.91 (1.6-2.2)	1.82 (1.3-2.3)	3.65 (2.6-5.0)	1.4×10^{-15}	3.4×10^{-15}	2.0×10^{-15}	1.9×10^{-12}	0.49 (1)	0.34	0.33	0.33
rs11978267	*IKZF1*	7	A/G*	1.69 (1.4-1.9)	1.88 (1.3-2.6)	2.99 (2.1-4.8)	8.8×10^{-11}	3.1×10^{-10}	4.5×10^{-12}	9.8×10^{-9}	0.39 (0.40)	0.27	0.26	0.27
rs10994982	*ARID5B*	10	G/A*	1.61 (1.3-1.9)	1.58 (1.1-2.2)	2.59 (1.8-3.6)	5.7×10^{-9}	8.5×10^{-9}	3.7×10^{-9}	5.5×10^{-7}	0.61 (1)	0.49	0.42	0.5
rs2089222	*KRTHB5*	12	G/A*	2.26 (1.6-3.0)	2.13 (1.5-2.9)	8.38 (2.4-28)	8.4×10^{-8}	1.3×10^{-8}	1.8×10^{-4}	4.6×10^{-4}	0.07 (0.39)	0.03	0.04	0.04
rs11155133		6	A/G*	3.62 (2.1-6.0)	3.62 (2.1-6.0)	NA	3.0×10^{-7}	1.3×10^{-7}	2.7×10^{-4}	1.4×10^{-4}	0.03 (1)	0.01	0.01	0.01
rs2191566	*ZNF230*	19	T/G*	1.52 (1.2-1.7)	1.18 (0.8-1.6)	2.07 (1.4-3.0)	3.5×10^{-7}	3.2×10^{-7}	8.0×10^{-6}	1.3×10^{-6}	0.37 (0.19)	0.28	0.29	0.27
rs2242041	*DDC*	7	C/G*	1.72 (1.3-2.1)	1.90 (1.4-2.4)	1.58 (0.5-4.3)	9.9×10^{-7}	4.0×10^{-6}	2.1×10^{-8}	1.4×10^{-6}	0.15 (0.19)	0.1	0.08	0.09
rs6509133	*ZNF230*	19	T/C*	1.48 (1.2-1.7)	1.19 (0.8-1.7)	2.01 (1.3-2.9)	1.4×10^{-6}	8.9×10^{-7}	5.7×10^{-5}	2.9×10^{-5}	0.37 (0.28)	0.28	0.29	0.28
rs7554607	*RYR2*	1	G/A*	1.49 (1.2-1.7)	1.74 (1.2-2.5)	2.39 (1.6-3.5)	1.6×10^{-6}	2.6×10^{-6}	1.1×10^{-6}	1.9×10^{-6}	0.65 (0.54)	0.56	0.55	0.55
rs12621643	*KCNE4*	2	G/T*	1.48 (1.2-1.7)	1.69 (1.2-2.3)	2.33 (1.6-3.3)	2.8×10^{-6}	1.3×10^{-6}	1.4×10^{-4}	2.9×10^{-4}	0.37 (0.63)	0.28	0.3	0.3
rs2167364	*DDC*	7	T/C*	1.46 (1.2-1.7)	1.53 (1.1-1.9)	2.10 (1.4-2.9)	2.8×10^{-6}	1.1×10^{-5}	1.8×10^{-8}	2.9×10^{-6}	0.41 (0.65)	0.33	0.3	0.31
rs10873876	*SIAT7C*	1	C/T*	1.55 (1.2-1.8)	1.55 (1.2-1.9)	2.45 (1.4-4.1)	4.1×10^{-6}	1.3×10^{-6}	6.5×10^{-4}	1.6×10^{-3}	0.22 (0.75)	0.15	0.16	0.16
rs9290663	*KCNMB2*	3	A/T*	1.58 (1.2-1.9)	1.63 (1.2-2.0)	2.24 (1.2-4.1)	6.1×10^{-6}	4.3×10^{-6}	1.0×10^{-4}	3.4×10^{-4}	0.2 (0.86)	0.13	0.14	0.14
rs6428370		1	A/G*	1.43 (1.2-1.6)	1.51 (1.1-1.9)	1.99 (1.4-2.8)	7.1×10^{-6}	1.6×10^{-5}	1.3×10^{-6}	9.0×10^{-6}	0.4 (1)	0.32	0.3	0.3
rs1881797	*OR2C3*	1	T/C*	1.52 (1.2-1.8)	1.51 (0.9-2.4)	2.30 (1.4-3.7)	7.3×10^{-6}	2.4×10^{-5}	9.5×10^{-7}	1.6×10^{-7}	0.24 (0.88)	0.18	0.16	0.15
rs10849033	*C12orf5*	12	A/G*	2.55 (1.6-3.8)	2.62 (1.7-3.9)	NA	8.5×10^{-6}	3.9×10^{-6}	9.4×10^{-4}	1.6×10^{-3}	0.04 (0.41)	0.02	0.02	0.02

dbSNP ID	Gene name	Chr	Alleles (A/B)[a]	Allelic OR (95% CI)[b]	Genomic OR (95% CI)[c]		P value[d]				Allele Frequency[e]			
					A/B-A/A	B/B-A/A	ALL vs. combined control group	ALL vs. WTCCC	ALL vs. GAIN: Schiz.	ALL vs. GAIN: bipolar	ALL (HWE P value)	WTCCC	Schiz	Bipolar
rs1879352		18	T/C*	1.53 (1.2-1.8)	1.44 (0.8-2.4)	2.25 (1.3-3.8)	8.7×10^{-6}	1.0×10^{-5}	3.0×10^{-5}	4.9×10^{-5}	0.23 (0.88)	0.16	0.16	0.16
rs563507	*PARD3*	10	G/A*	2.0 (1.4-2.7)	NA	NA	9.1×10^{-6}	5.0×10^{-6}	2.6×10^{-4}	2.4×10^{-3}	0.08 (0.71)	0.04	0.04	0.05

[a] Asterisk denotes the B allele ("risk allele"), which had a higher frequency in patients with ALL than in the non-ALL control groups.

[b] Odds ratio for the likelihood of ALL among individuals carrying an additional copy of the B allele (vs. A allele) at each SNP listed. For example, individuals carrying an additional copy of the B allele (encoding cytosine) at the rs10821936 SNP locus were 1.91 times as likely as others to belong to the ALL group. 95% confidence intervals are shown in parenthesis.

[c] Odds ratio for the risk of ALL between the specified genotypes. NA, not applicable (odds ratio could not be estimated).

[d] Logistic regression analysis comparing allele frequency at the listed SNP locus between patients with ALL in the discovery cohort (n=317) and the combined non-ALL control group (n=17,958) and between patients with ALL in the discovery cohort and each of the non-ALL control groups: the Wellcome Trust Case Control Consortium (WTCCC) group, (n=14,311) and the schizophrenia (Schiz.) (n=2,601) and bipolar (n=1,046) study groups from the Genetic Association Informative Network.

[e] The allele frequency shown is listed for the B allele for each polymorphism in each population. Schiz., schizophrenia. In parentheses, the P value indicating the deviation of the observed allele frequencies for the discovery ALL population compared to that expected based on Hardy Weinberg equilibrium is shown.

Table 2

Germline SNPs whose allele frequency differed among four ALL subtypes.

dbSNP ID	Gene name	Chr	Alleles (A/B)[a]	Overall P value[b]	B allele frequency[c]					Subtype analysis				
					B-other	B-hyperdiploid	T-cell ALL	t(12;21)/ *ETV6-RUNX1*	Non-ALL Control	ALL subtype[d]	Allelic OR; 95% CI[e]	P value[f]	Genotypic OR; 95% CI[g]	
													*/-vs. -/-	*/* vs. -/-
rs10821936	*ARID5B*	10	T/C*	0.0001	0.44 (0.41)	0.61 (0.64)	0.38 (0.42)	0.42 (0.45)	0.33	B-hyperdiploid	**2.17; 1.5-3.1 (2.45; 1.3-4.4)**	**1.62×10^{-5} (0.003)**	2.26; 1.3-4.0 (2.04; 0.8-5.1)	4.65; 2.3-9.5 (6.56; 1.8-24)
rs10994982	*ARID5B*	10	G/A*	0.016	0.57 (0.43)	0.69 (0.74)	0.52 (0.54)	0.6 (0.53)	0.49	B-hyperdiploid	**1.72; 1.2-2.4 (2.86; 1.5-5.3)**	**0.003 (0.0008)**	1.73; 1.0-2.9 (3.08; 1.3-7.4)	2.94; 1.3-6.5 (7.62; 1.9-30)
rs11978267	*IKZF1*	7	A*/G	0.001	0.44 (0.29)	0.41 (0.38)	0.21 (0.17)	0.36 (0.23)	0.27	T-cell ALL	2.63; 1.5-4.5 (2.27; 1.0-5.2)	0.001 (0.05)	2.16; 1.1-4.3 (1.75; 0.7-4.5)	13.68; 1.8-106 (NA)
rs2167364	*DDC*	7	T*/C	0.008	0.46 (0.37)	0.43 (0.36)	0.26 (0.19)	0.41 (0.30)	0.31	T-cell ALL	2.32; 1.4-4.0 (2.43; 1.1-5.6)	0.002 (0.03)	1.79; 0.9-3.5 (2.62; 1.0-7.1)	13.2; 1.7-103 (4.68; 0.6-40)
rs2242041	*DDC*	7	C*/G	0.041	0.16 (0.14)	0.18 (0.13)	0.07 (0.02)	0.13 (0.05)	0.09	T-cell ALL	2.86; 1.2-6.7 (6.67; 0.8-50)	0.02 (0.07)	2.74; 1.1-6.8 (6.19; 0.8-49)	NA (NA)
rs1881797	*OR2C3*	1	C*/T	0.05	0.29 (0.18)	0.25 (0.18)	0.22 (0.12)	0.14 (0.25)	0.16	t(12;21)/ *ETV6-RUNX1*	2.08; 1.1-3.8 (1.72; 0.7-4.0)	0.021 (0.21)	NA (11.67; 1.0-142)	1.70; 0.8-3.4 (11.0; 0.8-145)

[a] A sterisks denote the allele associated with the indicated ALL subtype. For example, the C allele at the rs10821936 polymorphism is associated with B-hyperdiploid ALL; the A allele at the rs11978267 polymorphism is associated with T-cell ALL.

[b] Multinomial log linear comparison of the allele frequency of each SNP among the four ALL subtypes in the discovery cohort.

[c] Frequency of the B allele within each ALL subtype in the discovery and validation (in parentheses) cohorts. The "non-ALL control" frequency is the mean value from the three non-ALL control groups.

[d] The ALL subtype most strongly distinguished by the indicated polymorphism.

[e] Odds ratio for the indicated ALL subtype vs. all other subtypes among individuals carrying an additional copy of the subtype-associated allele (marked with asterisk) in the discovery and validation (in parentheses) cohorts. For example, individuals carrying an additional copy of the C allele at the rs10821936 polymorphism are 2.17 times more likely to have B-hyperdiploid ALL than all other ALL subtypes. 95% confidence intervals are shown.

[f] Logistic regression analysis comparing allele frequency of the listed polymorphism in the indicated ALL subtype vs. all other ALL subtypes combined. Polymorphisms listed in bold were significantly associated with the indicated ALL subtype in both the discovery and validation groups after correction for multiple testing.

[g] Odds ratio for the indicated ALL subtype in study subjects with the specified genotypes. */* represents the hom ozygous genotype for the subtype-associated allele. */- represents the heterozygous genotype for the subtype-associated allele. -/- represents the homozygous genotype for the non subtype-associated allele. NA, odds ratio cannot be estimated.

Nat Genet. Author manuscript; available in PMC 2010 March 1.

Table 3

Absolute genotype count and ***allele frequency*** (bold italics) for the risk alleles *ARID5B* in SNPs in ALL subtypes and non-ALL controls.

		Discovery cohort (n=317)				Validation cohort (n=124)				Non-ALL control groups		
***ARID5B* SNP**	**SNP genotype**	**B-other**	**B-hyperdiploid**	**T-cell ALL**	**t(12;21); *ETV6-RUNX1***	**B-other**	**B-hyperdiploid**	**T-cell ALL**	**t(12;21); *ETV6-RUNX1***	**WTCCC**	**Bipolar**	**Schiz**
rs10821936	TT	37	16	16	14	13	4	10	7	6207	453	1160
	TC	61	52	21	24	26	18	8	8	6508	488	1170
	CC	23	40	6	7	5	14	6	5	1581	105	271
	C allele frequency	***0.44***	***0.61***	***0.38***	***0.42***	***0.41***	***0.64***	***0.42***	***0.45***	***0.34***	***0.33***	***0.33***
rs10994982	GG	22	10	9	7	14	3	6	4	3687	256	672
	GA	58	46	23	22	21	13	10	11	7093	544	1336
	AA	39	51	11	16	8	20	8	5	3490	246	592
	A allele frequency	***0.57***	***0.69***	***0.52***	***0.60***	***0.43***	***0.74***	***0.54***	***0.53***	***0.49***	***0.50***	***0.49***