LYMPHOID NEOPLASIA

A genome-wide association study of susceptibility to acute lymphoblastic leukemia in adolescents and young adults

Virginia Perez-Andreu,1 Kathryn G. Roberts,2 Heng Xu,1,3 Colton Smith,1 Hui Zhang,1 Wenjian Yang,1 Richard C. Harvey,4 Debbie Payne-Turner,5 Meenakshi Devidas,6 I-Ming Cheng,6 William L. Carroll,8 Nyla A. Heerema,8 Andrew J. Carroll,9 Elizabeth A. Raetz,10 Julie M. Gastier-Foster,11 Guido Marcucci,12 Clara D. Bloomfield,12 Krzysztof Mrózek,12 Jessica Kohlschmidt,12,13 Wendy Stock,14 Steven M. Kornblau,15 Marina Konopleva,16 Elisabeth Paietta,17 Jacob M. Rowe,18 Selina M. Luger,19 Martin S. Tallman,20 Michael Dean,21 Esteban G. Burchard,22 Dara G. Torgerson,22 Feng Yue,23 Yanli Wang,23 Ching-Hon Pui,24 Sima Jeha,24 Mary V. Relling,1 William E. Evans,1 Daniela S. Gerhard,25 Mignon L. Loh,26 Cheryl L. Willman,6 Stephen P. Hunger,27 Charles G. Mullighan,26 and Jun J. Yang1

1Department of Pharmaceutical Sciences and 2Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN; 3State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, West China Medical School, Sichuan University, Chengdu, China; 4University of New Mexico Cancer Center, Albuquerque, NM; 5Department of Biostatistics, Colleges of Medicine, Public Health & Health Professions, University of Florida, Gainesville, FL; 6Cancer Research and Treatment Center, University of New Mexico, Albuquerque, NM; 7Laura and Isaac Perlmutter Cancer Center, New York University Langone Medical Center, New York, NY; 8Department of Pathology, College of Medicine, The Ohio State University, Columbus, OH; 9Department of Genetics, University of Alabama at Birmingham, Birmingham, AL; 10Huntsman Cancer Institute, The University of Utah, Salt Lake City, UT; 11The Ohio State University, Nationwide Children's Hospital, Columbus, OH; 12Comprehensive Cancer Center, The Ohio State University, Columbus, OH; 13Alliance of Clinical Trials in Oncology Statistics and Data Center, Mayo Clinic, Rochester; 14Department of Medicine, Section of Hematology/Oncology, University of Chicago, Chicago, IL; 15Department of Molecular Hematology & Therapy, Department of Leukemia, The University of Texas MD Anderson Cancer Center, Houston, TX; 16Office of Cancer Genomics, National Cancer Institute, Frederick, MD; 17Department of Bioengineering & Therapeutic Science and Department of Medicine, Pennsylvania State University College of Medicine, Hershey, PA; 18Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN; 19Department of Biostatistics, Colleges of Medicine, Public Health & Health Professions, University of Florida, Gainesville, FL; 20Leukemia Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY; 21Laboratory of Experimental Immunology, National Cancer Institute, Frederick, MD; 22Department of Bioengineering & Therapeutic Science and Department of Medicine, University of California, San Francisco, San Francisco, CA; 23Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey, PA; 24Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN; 25Office of Cancer Genomics, National Cancer Institute, Bethesda, MD; 26Department of Pediatrics and the Helen Diller Family Cancer Center, University of California, San Francisco, San Francisco, CA; and 27School of Medicine and Children's Hospital, University of Colorado School of Medicine, Aurora, CO

Key Points

- In this first ALL GWAS in AYAs, we determined that inherited GATA3 variants strongly influence ALL susceptibility in this age group.
- These findings revealed similarities and differences in the genetic basis of ALL susceptibility between young children and AYAs.

Acute lymphoblastic leukemia (ALL) in adolescents and young adults (AYA) is characterized by distinct presenting features and inferior prognosis compared with pediatric ALL. We performed a genome-wide association study (GWAS) to comprehensively identify inherited genetic variants associated with susceptibility to AYA ALL. In the discovery GWAS, we compared genotype frequency at 635,297 single nucleotide polymorphisms (SNPs) in 308 AYA ALL cases and 6,661 non-ALL controls by using a logistic regression model with genetic ancestry as a covariate. SNPs that reached $P \leq 5 \times 10^{-8}$ in GWAS were tested in an independent cohort of 162 AYA ALL cases and 5,755 non-ALL controls. We identified a single genome-wide significant susceptibility locus in GATA3: rs3824662, odds ratio (OR), 1.77 ($P = 2.8 \times 10^{-10}$) and rs3781093, OR, 1.73 ($P = 3.2 \times 10^{-9}$). These findings were validated in the replication cohort. The risk allele at rs3824662 was most frequent in Philadelphia chromosome (Ph)-like ALL but also conferred susceptibility to non-Ph-like ALL in AYAs. In 1,827 non-selected ALL cases, the risk allele frequency at this SNP was positively correlated with age at diagnosis ($P = 6.29 \times 10^{-11}$). Our results from this first GWAS of AYA ALL susceptibility point to unique biology underlying leukemogenesis and potentially distinct disease etiology by age group. (Blood. 2015;125(4):680-686)

Introduction

Cancer survival rates have been steadily increasing in the United States across age groups except for in adolescents and young adults (AYA; age 16 to 39 years), partly because of the persisting inferior treatment response in hematologic malignancies. Particularly with acute lymphoblastic leukemia (ALL), age as a continuous variable is negatively correlated with prognosis in spite of risk-adapted combination chemotherapy. In an analysis of 21,626 ALL cases diagnosed between 1990 and 2005 and treated on Children’s Oncology Group (COG) frontline protocols, survival rates decreased significantly with increasing age at diagnosis regardless of treatment era (eg, 94.1% for...
age 1 to 10, 84.7% for age 10 to 15, and 75.9% for age 15 to 22 years in the 2000-2005 cohort. Although pediatric-based treatment regimens have been tested in AYA populations and have resulted in improved survival, the gap in treatment outcome between age groups persists, and ALL remains one of the leading causes of cancer-related deaths in the AYA population.4,5

The inferior prognosis of AYA is likely to be multifactorial and includes socioeconomic factors, medication adherence, clinical trial enrollment and, importantly, age-related differences in ALL tumor and host biology.6 For example, as age increases, there is a progressive rise in prevalence of ALL genetic subtypes with poor prognosis such as Philadelphia chromosome–positive (Ph+)7,7 Ph-like8,8 or intrachromosomal amplification of chromosome 21,9 whereas subtypes with favorable outcome (high hyperdiploidy10 and ETV6-RUNX111) become less common. These comparisons are informative but are also limited because they are primarily driven by ALL features discovered in children and/or older adults. As a result, differences in tumor biology between AYA and childhood ALL may have been underestimated, and genomic profiling studies focusing on the AYAs are likely to reveal novel molecular features unique to this population.

Inherited genetic variations can strongly influence both the susceptibility to ALL12-15 and treatment outcomes.16-20 For example, genome-wide association studies (GWASs) have identified germline genetic variants at ARID5B, IKZF1, CEBPE, PIP4K2A, and CDKN2A/CDKN2B loci with substantial cumulative effects on ALL disease risk in children. These ALL susceptibility genes are involved in lymphoid cell development, cell cycle control, and tumor suppression, collectively affecting leukemogenesis. Although one’s inherited genetic variants remain unchanged over a lifetime, it is possible that the effects of these susceptibility variants vary by age, thus contributing to the age-related differences in ALL incidence and subtype. In fact, when we examined the risk of ALL conferred by the ARID5B variant, there was a clear trend of diminishing effects with increasing age (e.g., allelic odds ratio of 2.01, 1.8, and 1.48 in children younger than age 5, 5 to 10, and older than 10 years, respectively).15 However, germline variants related to ALL risk in the AYA population have not been comprehensively examined.

To better understand the potential unique leukemogenesis in AYAs, we conducted the first GWAS to systematically interrogate germline single nucleotide polymorphisms (SNPs) for their contribution to ALL risk in this age group.

Methods

Study design and patients

In the discovery GWAS, the ALL cases consisted of 209 adolescents (median age, 17.4 years; range, 16 to 21 years) and 99 young adults (median age, 24.3 years; range, 21 to 39 years) with newly diagnosed B-cell ALL who were treated on the Children’s Oncology Group (COG; N = 202),21 the Alliance-Cancer and Leukemia Group B (N = 56),22 Eastern Cooperative Oncology Group E2993 (N = 29),23 MD Anderson Cancer Center (N = 11)24-27, and St. Jude Children’s Research Hospital (N = 10) trials.28 The subjects were chosen on the basis of the availability of germline DNA, which was extracted from peripheral blood samples during clinical remission (<5% blasts cells in bone marrow). A total of 6,661 unrelated subjects from the Multi-Ethnic Study of Atherosclerosis (MESA) cohort (dbGaP phs0000209.v9) were considered as non-ALL control subjects because the prevalence of adult survivors of childhood ALL is extremely low.14

For the replication analyses, 162 children with ALL age 16 to 21 years from the COG P9900 protocols (COG P9905 [NCT00005596], COG P9904 [NCT0000558515], COG P9906 [NCT0000560331], and COG AALL0232 [NCT0007572531]) were included. A set of 5755 unrelated non-ALL controls were included in the replication analysis: 1228 African Americans (AAs) from the AIDS Linked to Intravenous Experience (ALIVE) cohort,31 880 Hispanic Americans from the Genetics of Asthma in Latino Americans (GALA) study,32 and 3647 European Americans (EAs) from the Genetic Association Informative Network (GAIN) schizophrenia cohort (dbGaP phs000021.v3.p2)15 and GAIN bipolar cohort (phs000017.v3.p114, Figure 1).

The clinical trials were approved by local institutional review boards, and informed consent for trial enrollment and banking of specimens for future research was obtained from parents, guardians, or patients, as appropriate. This study was approved by the St. Jude Children’s Research Hospital institutional review board.

Genotyping and quality control

Genome-wide SNP genotyping was performed by using the Affymetrix Human SNP Array 6.0 for ALL cases in the discovery GWAS, those in the COG P9905, P9904, and AALL0232 cohorts, and for all non-ALL controls (dbGaP MESA, ALIVE, GAIN, and GALA). Genotype calls (coded as 0, 1, and 2 for AA, AB,
and BB genotypes) were determined by the Birdseed v2 (Affymetrix SNP 6.0) algorithm. Samples for which genotypes were ascertainment for less than 95% of SNPs on the array were deemed to have failed and were excluded from the analyses. For the ALL cases in the COG P9006 trial, genome-wide SNP genotyping was performed by using Affymetrix Human SNP Array 500K, and GATA3 SNPs were genotyped by polymerase chain reaction and Sanger sequencing, as described previously. We did not observe evidence of potential genotyping errors in the germline DNA because of tumor cell contamination (data not shown).

Prior to GWAS, SNPs were subjected to a series of quality control steps (supplemental Figure 1, available on the Blood Web site). First, we filtered SNPs on the basis of minor allele frequency (MAF) and SNP call rate: for SNPs with an MAF of 1% to 3%, we excluded those with a call rate <99%; for SNPs with an MAF of 3% to 5%, we excluded those with a call rate <98%; for SNPs with an MAF of >5%, we excluded those with a call rate <95%. An additional filtering step was applied in the GWAS involving non-ALL controls: we removed SNPs for which genotype frequencies differed significantly among control groups (ie, dbGaP MESA vs HapMap unrelated CEU or dbGaP MESA vs the GAIN bipolar cohort [dbGaP phs000017.v3]). P < 10−6 by χ2 test, and the comparison was restricted to EAs. Finally, those SNPs deviating from Hardy-Weinberg equilibrium (P < .01 in EA cases or controls) were also excluded from the analysis. After quality control filters were applied, 635 297 SNPs were included in the GWAS.

Genetic ancestry and population structure

Genetic ancestry was determined by using STRUCTURE (version 2.2.3),37 based on genotypes at 30 000 SNPs randomly selected from the Affymetrix SNP arrays. HapMap samples from descendants of Northern Europeans (CEU; N = 60), West Africans (YRI; N = 60), East Asians (CHB/JPT; N = 90), and Native Americans (NA; N = 105)38 references were used to represent European, African, and Native American ancestries, respectively. We assumed that these four ancestries summed to 100% in each genotyped individual. EAs, AAs, and Hispanic Americans were genotyped by polymerase chain reaction and Sanger sequencing, as described previously. We did not observe evidence of potential genotyping errors in the germline DNA because of tumor cell contamination (data not shown).

Prior to GWAS, SNPs were subjected to a series of quality control steps (supplemental Figure 1, available on the Blood Web site). First, we filtered SNPs on the basis of minor allele frequency (MAF) and SNP call rate: for SNPs with an MAF of 1% to 3%, we excluded those with a call rate <99%; for SNPs with an MAF of 3% to 5%, we excluded those with a call rate <98%; for SNPs with an MAF of >5%, we excluded those with a call rate <95%. An additional filtering step was applied in the GWAS involving non-ALL controls: we removed SNPs for which genotype frequencies differed significantly among control groups (ie, dbGaP MESA vs HapMap unrelated CEU or dbGaP MESA vs the GAIN bipolar cohort [dbGaP phs000017.v3]), P < 10−6 by χ2 test, and the comparison was restricted to EAs. Finally, those SNPs deviating from Hardy-Weinberg equilibrium (P < .01 in EA cases or controls) were also excluded from the analysis. After quality control filters were applied, 635 297 SNPs were included in the GWAS.

Genetic ancestry was determined by using STRUCTURE (version 2.2.3), based on genotypes at 30 000 SNPs randomly selected from the Affymetrix SNP arrays. HapMap samples from descendants of Northern Europeans (CEU; N = 60), West Africans (YRI; N = 60), East Asians (CHB/JPT; N = 90), and Native Americans (NA; N = 105) references were used to represent European, African, and Native American ancestries, respectively. We assumed that these four ancestries summed to 100% in each genotyped individual. EAs, AAs, and Hispanic Americans were genotyped by polymerase chain reaction and Sanger sequencing, as described previously. We did not observe evidence of potential genotyping errors in the germline DNA because of tumor cell contamination (data not shown).

Prior to GWAS, SNPs were subjected to a series of quality control steps (supplemental Figure 1, available on the Blood Web site). First, we filtered SNPs on the basis of minor allele frequency (MAF) and SNP call rate: for SNPs with an MAF of 1% to 3%, we excluded those with a call rate <99%; for SNPs with an MAF of 3% to 5%, we excluded those with a call rate <98%; for SNPs with an MAF of >5%, we excluded those with a call rate <95%. An additional filtering step was applied in the GWAS involving non-ALL controls: we removed SNPs for which genotype frequencies differed significantly among control groups (ie, dbGaP MESA vs HapMap unrelated CEU or dbGaP MESA vs the GAIN bipolar cohort [dbGaP phs000017.v3]), P < 10−6 by χ2 test, and the comparison was restricted to EAs. Finally, those SNPs deviating from Hardy-Weinberg equilibrium (P < .01 in EA cases or controls) were also excluded from the analysis. After quality control filters were applied, 635 297 SNPs were included in the GWAS.

Results

AYA ALL GWAS

In the discovery GWAS, we compared genotype frequency at 635 297 SNPs between 308 AYA ALL cases and 6,661 non-ALL controls (Figure 1). After adjusting for genetic ancestry, only two SNPs at 10p14 within the GATA3 gene reached genome-wide significance: rs3824662 (odds ratio [OR], 1.77; 95% confidence interval [CI], 1.48 to 2.12; P = 2.84 × 10−10) and rs3781093 (OR, 1.73; 95% CI, 1.44 to 2.08; P = 3.20 × 10−5; Table 1 and Figure 2). These two SNPs were in strong linkage disequilibrium (r2 = 0.94; D' = 1 in HapMap CEU; supplemental Figure 4), representing a single susceptibility locus. The A allele at rs3824662 was significantly overrepresented in ALL cases compared with controls (35% vs 20%) and was consistent across race/ethnicity (ie, EAs, 30% vs 17% [P = 1.09 × 10−3]; Hispanics, 50% vs 33% [P = .0008]; and AAs, 20% vs 10% [P = .07]; Figure 3A).

rs3781093 was significantly associated with ALL risk in EAs and Hispanics, but not in individuals of African descent in whom it was...
no longer in linkage disequilibrium ($r^2 = 0.006$; $D' = 0.16$) with rs3824662 (supplemental Figure 5A).

To validate the association signals at these GATA3 SNPs, we tested an independent set of 162 AYA ALL cases enrolled in COG P9900 and ALL0232 protocols and an additional 5,755 non-ALL controls. In the replication analysis, risk alleles at both GATA3 SNPs were consistently overrepresented in AYA ALL cases compared with non-ALL controls: rs3824662 (OR, 2.21; 95% CI, 1.72 to 2.83; $P = 1.52 \times 10^{-10}$) and rs3781093 (OR, 1.96; 95% CI, 1.52 to 2.54; $P = 1.03 \times 10^{-7}$; Table 1, Figure 3B, and supplemental Figure 5B). rs3824662 was validated across race/ethnicity in the replication group (ie, EAs, 35% vs 18% [P = 2.0 x 10^{-7}]; Hispanics, 55% vs 39% [P = .005]; and AAs, 13% vs 9% [P = .035]; Figure 3B). In contrast, rs3781093 was significant in EAs and Hispanics but not in AAs (supplemental Figure 5B).

We also examined the association signals in AYAs for susceptibility loci previously identified in pediatric populations (supplemental Table 1). ARID5B, IKZF1, and PIP4K2A variants were nominally significant in AYAs in the discovery GWAS and/or in the replication analyses. In contrast, CEBPE and CDKN2A/CDKN2B were not associated with ALL risk in AYAs in either discovery or replication cohorts. These results imply both similarities and differences in genetic predisposition to ALL between children and AYAs.15

**GATA3 SNP rs3824662 and ALL subtypes in AYAs**

We further analyzed the association of the GATA3 SNP rs3824662 with somatic ALL genomic abnormalities. Among the AYA ALL cases in the discovery cohort, the risk allele at rs3824662 was underrepresented among hyperdiploid ALL cases (22% vs 37%; $P = .03$; Figure 4 and supplemental Table 2), with a similar trend for TCF3-PBX1 and ETV6-RUNX1 ALL albeit not statistically significant. In contrast, the ALL risk allele frequency of rs3824662 was higher in AYA ALL cases with the Ph-like gene expression profile than in those without this signature (48% vs 32%; $P = .02$; Figure 4 and supplemental Table 2). This was consistent with our prior reports of GATA3 as a susceptibility gene for Ph-like ALL,36 although there was no overlap in cases included in the current AYA ALL GWAS and those in our previous Ph-like ALL GWAS.36 Within Ph-like ALL, there was a trend with A allele further enriched in cases involving CRLF2 rearrangements ($P = .06$; Figure 4).
Importantly, even after excluding Ph-like ALL cases, the risk allele at rs3824664 was still more common in AYA ALL cases compared with non-ALL controls (rs3824662: OR, 1.56 [95% CI, 1.25 to 1.96; \( P = 8.13 \times 10^{-3} \); rs3781093: OR, 1.53 [95% CI, 1.21 to 1.92; \( P = .0002 \)]; supplemental Figure 6). This suggested that the influence of the GATA3 variant on ALL susceptibility in AYAs extends beyond the predisposition to Ph-like subtype.

GATA3 SNP rs3824662 and age at ALL diagnosis

Finally, we examined the distribution of the GATA3 SNP genotype by age at diagnosis in a cohort of largely unselected patients enrolled on the COG P9900 protocols (N = 1,827, age 0.1 to 21 years). When we divided patients into four consecutive age groups (<5, 5 to 10, 10 to 15, and >15 years), we observed a clear progressive increase in the risk allele frequency at rs3824662 (\( P = 6.29 \times 10^{-11} \); Figure 5) with increasing allelic ORs (ie, relative risk of ALL conferred by each copy of the A allele at rs3824662; Figure 5, inset plot): 0.96 (95% CI, 0.85 to 1.09), 1.26 (95% CI, 1.08 to 1.48), 1.48 (95% CI, 1.19 to 1.84), and 2.40 (95% CI, 1.81 to 3.19). Similar correlation between GATA3 genotype frequency and age was evident irrespective of genetic ancestry, but the GATA3 risk allele was markedly more common in Hispanics (ie, individuals with high NA genetic ancestry; Figure 5). To examine whether the association with age is confounded by ALL genetic subtype, we compared rs3824662 allele frequency by age in the COG P9900 protocols after stratifying ALL cases into TCF3-PBX1, ETV6-RUNX1, high hyperdiploid, MLL-rearranged, and B-other. There was a trend for the risk allele at this SNP to be more frequent in patients older than age 16 years relative to those younger than age 16 years in 5 subtypes examined, although with a limited sample size (supplemental Figure 7). This suggests that GATA3 germline variants confer a general ALL disease risk in AYAs. In contrast, the frequency of ALL risk variant in ARID5B (rs10821936) decreased progressively with increasing age at diagnosis in the COG P9900 cohort (\( P = .006 \)), whereas PIP4K2A, CDKN2A/2B, IKZF1, and CEBPE variants were not related to age (\( P > .05 \); data not shown).

Discussion

Because ALL is the most common cancer in children, previous susceptibility GWAS studies understandably focused on pediatric populations. We hypothesized that ALL in AYAs has distinct tumor biology and genetic etiology, which potentially contribute to the disparities in treatment.
outcomes by age. To this end, we performed the first GWAS of ALL susceptibility specifically in the AYA population and identified a single genome-wide significant risk locus within the GATA3 gene on 10p14.

The susceptibility to ALL varies substantially by age. ALL risk first peaks between 2 and 5 years after birth, followed by gradual decrease into adulthood, but rises again in older individuals (older than age 70 years), suggesting that differential combinations of environmental and genetic factors contribute to leukemogenesis at different ages. For example, it has been hypothesized that infection (and supposedly acquired immunity) may ameliorate susceptibility to ALL in young children,[41,42] which may not be important in ALL that occurs later in life. Similarly, the in utero occurrence of genomic lesions is characteristic in many (if not most) pediatric ALL cases,[43,44] whereas such early origin of presumed initiating events may not be evident in AYAs ALL. Age-dependent differences in lymphocyte development and function are well documented in human and mouse systems,[45] and rapid growth of hematopoietic cells may render them particularly susceptible to oncogenic assaults.[46] Thus, it can be postulated that specific ALL susceptibility genes are required during a particular stage of hematopoietic development and preferentially influence ALL risk within a certain age range. For example, loss of Arid5b in mice resulted in reduction of lymphoid cells in bone marrow within 3 weeks after birth, but the effect became blunted by 6 weeks.[47] In fact, germline ARID5B variants also exhibited increasing influence on ALL predisposition in children as age decreases.[48]

GATA3 encodes for a transcription factor critical for lymphoid lineage commitment and early T-cell differentiation,[49] and loss-of-function somatic mutations have been discovered in early T-cell precursor ALL.[49] Germine polymorphisms in GATA3, however, appear more important for B-cell malignancies.[50] We recently reported that rs3824662 was significantly associated with Ph-like ALL in children and risk of relapse.[36] A contemporaneous study by Migliorini et al[17] reported the same ALL susceptibility variant in GATA3 in children of European descent and associated it with relapse. Particularly of note, GATA3 risk variants also appeared enriched in older children, even within their predominantly pediatric cohort. The association of rs3824662 with ALL relapse[17,36] is in line with the negative prognosis by age and higher frequency of the GATA3 variant in AYAs with ALL. Nevertheless, it is unclear whether poor prognosis conferred by a GATA3 variant was driven by its association with a high-risk subtype (ie, Ph-like ALL), novel somatic genomic aberrations specific to AYA, and/or host biology related to antileukemic drug response. Interestingly, in AYA cases included in the discovery GWAS, the number of the risk allele at rs3824662 was significantly associated with GATA3 expression in ALL blasts (P = .02; supplemental Figure 8), consistent with our previous report in pediatric ALL cases of this variant functioning as a cis-acting regulatory element of GATA3 transcription.[36]

The overrepresentation of the GATA3 variant in AYAs is consistent with its association with Ph-like ALL[36] for which the frequency increases with age.[8] However, the risk variant at rs3824662 remained significantly associated with susceptibility to ALL cases without Ph-like expression pattern, suggesting the link to Ph-like ALL contributed only partly to the genome-wide significant association signal at rs3824662. In fact, the GATA3 risk allele tended to be more common in ALL patients age 16 years or older than in those age younger than 16 years consistently across different genetic subtypes, plausibly conferring a general ALL risk in AYAs. It remains unknown how the GATA3 variants influence the risk of developing ALL in older adults, including the elderly (>60 years). Future studies including this age group may provide insights on molecular etiology of ALL across the age spectrum. It is also noteworthy that MLL-rearranged cases had the second highest GATA3 risk variant frequency (Figure 4), although the number of patients was relatively small and the difference did not reach statistical significance. Future studies are warranted to comprehensively characterize potential interactions of germline GATA3 variants with somatic genomic lesions in ALL.

In conclusion, our GWAS identified inherited GATA3 genetic variants that strongly influence ALL susceptibility in adolescents and young adults, shedding new light on potential age-related differences in ALL biology and treatment outcome.

Acknowledgments

The authors thank the patients and parents who participated in the clinical trials included in this study and M. Shriver (Pennsylvania State University) for sharing SNP genotype data for the Native American references. Genome-wide genotyping of COG P9904/ P9905 samples was performed by the Center for Molecular Medicine with the generous financial support from the Jeffrey Pride Foundation and the National Childhood Cancer Foundation.

This work was supported by the National Institutes of Health, National Cancer Institute grants CA145707, CA156449, CA21765, CA236401, CA98543, CA114766, CA98413, CA140729, CA176063, and HHSN261200800001E, the National Institute of General Medical Sciences grant GM92666, in part by the intramural Program of the National Cancer Institute, by a Stand Up to Cancer Innovative Research Grant, and by the American Lebanese Syrian Associated Charities of St. Jude Children’s Research Hospital, by a St. Jude Children’s Research Hospital Academic Programs Special Fellowship and a Spanish Ministry of Education Fellowship Grant (V.P.-A.), by the American Society of Hematology Scholar Award and by the Order of St. Francis Foundation (J.J.Y.), and by a Leukemia and Lymphoma Society Fellow Award and Alex’s Lemonade Stand Foundation Young Investigator Award (K.G.R.). S.P.H. is the Ergen Family Chair in Pediatric Cancer, C.G.M. is a Pew Scholar in the Biomedical Sciences and a St. Baldrick’s Scholar, and H.Z. is a St. Baldrick’s International Scholar.

The study sponsors were not directly involved in the design of the study, the collection, analysis, and interpretation of the data, the writing of the manuscript, or the decision to submit the manuscript.

Authorship

Contribution: V.P.-A., K.G.R., H.X., C.G.M., and J.J.Y. conceived and designed the study; R.C.H., D.P.-T., I-M.C., W.L.C., N.A.H., A.J.C., E.A.R., J.M.G.-F., G.M., C.D.B., K.M., E.G.B., D.G.T., F.Y., Y.W., S.M.L., M.D., E.P., J.M.R., S.M.K., M.S.T., M.D., E.G.B., D.G.T., F.Y., Y.W., C.-H.P., S.J., M.V.R., W.E.E., D.S.G., M.L.L., S.P.H., and C.L.W. provided study materials or patients; V.P.-A., K.G.R., H.X., M.D., I-M.C., C.L.W., R.C.H., M.V.R., and W.E.E. contributed only partly to the genome-wide significant association signal at rs3824662. In fact, the GATA3 risk allele tended to be more common in ALL patients age 16 years or older than in those age younger than 16 years consistently across different genetic subtypes, plausibly conferring a general ALL risk in AYAs. It remains unknown how the GATA3 variants influence the risk of developing ALL in older adults, including the elderly (>60 years). Future studies including this age group may provide insights on molecular etiology of ALL across the age spectrum. It is also noteworthy that MLL-rearranged cases had the second highest GATA3 risk variant frequency (Figure 4), although the number of patients was relatively small and the difference did not reach statistical significance. Future studies are warranted to comprehensively characterize potential interactions of germline GATA3 variants with somatic genomic lesions in ALL.

In conclusion, our GWAS identified inherited GATA3 genetic variants that strongly influence ALL susceptibility in adolescents and young adults, shedding new light on potential age-related differences in ALL biology and treatment outcome.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Jun J. Yang, Department of Pharmaceutical Sciences, MS 313, St. Jude Children’s Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105-3678; e-mail: jun.yang@stjude.org; and Charles G. Mullighan, Department of Pathology, MS 342, St. Jude Children’s Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105-3678; e-mail:charles.mullighan@stjude.org.
References


A genome-wide association study of susceptibility to acute lymphoblastic leukemia in adolescents and young adults