Is There Etiologic Heterogeneity between Subtypes of Childhood Acute Lymphoblastic Leukemia? A Review of Variation in Risk by Subtype

Lindsay A. Williams1, Jun J. Yang2,3, Betsy A. Hirsch4,5, Erin L. Marcotte1,5, and Logan G. Spector1,5

Abstract

Although substantial advances in the identification of cytogenomic subtypes of childhood acute lymphoblastic leukemia (ALL) have been made in recent decades, epidemiologic research characterizing the etiologic heterogeneity of ALL by subtype has not kept pace. The purpose of this review is to summarize the current literature concerning subtype-specific epidemiologic risk factor associations with ALL subtype defined by immunophenotype (e.g., B-cell vs. T-cell) and cytogenomics (including gross chromosomal events characterized by recurring numerical and structural abnormalities, along with cryptic balanced rearrangements, and focal gene deletions). In case-control analyses investigating nongenetic risk factors, home paint exposure is associated with hyperdiploid, MLL-rearranged, and ETV6-RUNX1 subtypes, yet there are few differences in risk factor associations between T- and B-ALL. Although the association between maternal smoking and ALL overall has been null, maternal smoking is associated with an increasing number of gene deletions among cases. GWAS-identified variants in ARID5B have been the most extensively studied and are strongly associated with hyperdiploid B-ALL. GATA3 single nucleotide variant rs3824662 shows a strong association with Ph-like ALL (OR = 3.14). However, there have been relatively few population-based studies of adequate sample size to uncover risk factors that may define etiologic heterogeneity between and within the currently defined cytogenomic ALL subtypes.

Introduction

Cancer, regardless of age at diagnosis, is universally heterogeneous with variation in incidence by both immunophenotype and molecular subtype within tumor types. Studies of adult neoplasms including breast (1–5), colon (6–8), and prostate (9, 10), have shown that subtyping, whether molecular or histologic, leads to identification of etiologic heterogeneity between tumor subtypes that may inform targeted public health prevention strategies. For example, in breast cancer, there are robust findings across multiple study populations showing evidence of etiologic heterogeneity by reproductive risk factors such as age at menarche, first birth, and menopause, as well as combined menopausal hormone therapy use for lobular, compared with ductal carcinoma, and by molecular breast cancer subtype (1–5). Similarly, there is etiologic variation by histologic and molecular subtype for colorectal adenomas concerning combined hormone therapy use (6) and cigarette smoking (6–8). Finally, there is reported variability in risk of prostate tumors by TMPRSS2-ERG fusion status concerning age at diagnosis (9), height, and obesity (10). These examples highlight the presence of etiologic heterogeneity in adult breast, colon, and prostate tumors where once histology or molecular subtype are considered, differences in risk factor associations become clearer. Therefore, it is reasonable to hypothesize that studying variation in susceptibility to childhood acute lymphoblastic leukemia (ALL) by immunophenotype, and cytogenomic subtype, may similarly identify etiologic variation of this most common malignancy diagnosed in children aged 0–14 years (11). The purpose of this review is to summarize the current state of the literature regarding etiologic variation in ALL subtypes. We will discuss the literature on nongenetic and genetic risk factors for ALL overall and by subtype as defined by immunophenotype and cytogenomic subtype.

Descriptive Epidemiology of Childhood ALL

ALL is the most common childhood malignancy in the United States (US), with approximately 3,000 children and adolescents aged 0–19 years diagnosed annually (11). As shown in Fig. 1A, among children (aged 0–19 years) from the SEER 18 registries (2000–2014), there is notable variation in the age-specific incidence rates over the childhood and adolescent periods with the...
highest incidence rate for ALL between ages 2 and 4 years. Incidence rates of ALL continue to slowly increase necessitating the identification of modifiable risk factors (11, 12).

Even though 5-year overall survival rates of ALL have drastically improved since the 1970’s and are currently at approximately 90% (13), there are documented differences in survival by age at diagnosis, race, and subtype (13–15). Generally, once differences in survival and clinical outcome by disease subtype are reported, epidemiologic analyses are conducted on individual subtypes to identify etiologic differences (16). However, despite documented differences in prognosis and biology between pediatric ALL subtypes, there is a paucity of studies aimed at identifying unique risk factor profiles for these subtypes. This is likely due to older and registry-based studies lacking cytogenomic and molecular genetic data and/or inadequate sample sizes.

Biology of ALL

Immunophenotypes

ALL is composed of two main immunophenotypes that are identified by distinctive hematopoietic lineage markers: B- and T-ALL (17). Pediatric B-ALL is diagnosed in up to 85% of cases while T-ALL comprises the remaining 15% (18, 19) and is more common among males (20). There is racial/ethnic variation in the incidence of B- and T-ALL. Approximately 85% of white, 87% of Hispanic, 81% of Asian, and 75% of black children are diagnosed with B-ALL and the remaining proportion of children in each race/ethnicity are diagnosed with T-ALL (21–24), which has lower survival than B-ALL (13, 15).

Of the 15% of pediatric cases diagnosed with T-ALL, approximately 15% of these patients are diagnosed early T-precursor ALL (EPT-ALL; refs. 25–28). EPT-ALL cells have retained myeloid and stem cell characteristics making it a distinct entity from T-ALL (27, 29–31). The recent identification of effective therapies for ETP-ALL has shown it to be similar to T-ALL in outcome and prognosis (25, 26, 32). As observed for T-ALL, ETP-ALL is more common among males (ETP-ALL: 63% and T-ALL: 74%; ref. 25). The difference in etiology of ETP-ALL and T-ALL is unclear.

Cytogenomic subtypes

Cytogenomic subtypes refer to specific numerical and/or structural chromosomal abnormalities that are well documented to recur in different subgroups of ALL and are used in the clinic to risk-stratify patients. Cytogenomic subtyping usually occurs through visual chromosome assessment by use of FISH, karyotyping via G-bandning, and/or chromosomal microarray (14).

Figure 1B displays the overall and age-specific distributions of cytogenomic subtypes of B-ALL and the overall cytogenomic distribution for T-ALL (Fig. 1C; cytogenomics subtypes assumed to be mutually exclusive; refs. 20, 21, 23, 31, 33–36). Describing
the differences in clinical characteristics of cytogenetic subtypes is beyond the scope of this review and can be found in greater detail elsewhere (31, 36). The most marked variation in B-ALL incidence patterns is observed by age at diagnosis, which may also serve as a proxy for some cytogenomic subtypes. A prime example of this is observed among infant ALL, diagnosed in children less than 1 year of age, where 80% of cases have an MLL rearrangement (Fig. 2). Extending beyond infants, approximately 35% of cases with B-ALL aged 1–9 years are diagnosed with the hyperdiploid subtype and another 30% are diagnosed with ETV6-RUNX1 fusions (28).

Ploidy changes, the most common chromosomal abnormality in childhood B-ALL, are observed in approximately 35% cases (21, 28, 33–35, 37). Hyperdiploidy, the most common subtype, involves acquiring extra chromosomes including 4, 6, 10, 14, 17, 18, 21, and X chromosomes (23, 31, 38, 39). Chromosomal translocations, which often place transcription factors (e.g., ETV6, RUNX1) next to one another leading to production of novel functional proteins, dysregulation of gene expression, or activation of an oncogene (19, 31, 40–43), are the second most frequent gross chromosomal changes in ALL. In B-ALL, the ETV6-RUNX1 subtype, which typically arises from a cryptic t(12;21) fusion, that can be visualized by FISH, is diagnosed in approximately 30% of cases (28, 34, 44, 45). The second most common translocation, involving MLL rearrangement (MLLr; also referred to as KMT2A rearrangement) [t(4;11)(q21;q23)], is diagnosed in approximately 10% of cases overall, but 80% of infant B-ALL cases (20, 28, 34, 44–48). MLL is known to have more than 100 partner genes with which it can pair, and the specific pairing impacts prognosis (31).

There are a number of less frequently diagnosed cytogenomic subtypes involving gene fusions and the presence of dicentric chromosomes (28, 49), that vary by age at diagnosis as illustrated in Fig. 1B.

Figure 2.
Odds ratios (OR) and 95% confidence intervals (95% CI) for risk factors in association with childhood ALL overall and by immunophenotype and cytogenomic subtype presented by timing of exposure for preconception (A), pregnancy (B), birth (C), and childhood (D). Reference numbers in brackets.
While much work identifying cytogenomic subtypes has focused on the more common B-ALL immunophenotype of childhood ALL, cytogenomic subtypes have also been identified for T-ALL. T-ALL cases are rarely hyperdiploid and do not display the pattern of chromosomal gains seen in B-ALL, but can be triploid or tetraploid (31, 33). More often they have only 46 chromosomes, but harbor structural abnormalities and can be grouped into 4 main categories involving translocations of (i) TAL1 or LMO2; (ii) TLX1; (iii) TLX3; and (iv) HOX genes (31). T-ALL chromosomal translocations are characterized by transcription factor dysregulation through gene fusions (34). TAL1 fusions t(1;7)(p32;q35) and t(1;14)(p32;q11) are diagnosed in an estimated 40% of cases (26, 28, 34). In approximately 20% of cases, TLX3 [HOX11L2; t(5;14)(q35;32)] dysregulation is reported (26, 28, 31, 34, 50). The remaining fusions in Fig. 2 are diagnosed in 10% or less of pediatric cases each (26, 28, 34, 50). There is insufficient information on variation of these T-ALL subtypes by sex and race/ethnicity for the most part.

Molecular genomic abnormalities

Molecular genomic abnormalities, or genetic events that occur at subchromosomal levels, are detected by gene sequencing or expression patterns and have also been characterized for childhood ALL (28, 51). Most recently, the Philadelphia-like subtype [Ph-like, BCR-ABL1-like] B-ALL has been identified by gene expression patterns similar to that found in BCR-ABL1-positive B-ALL and has been further shown to have mutations affecting genes important in leukemogenesis such as IKZF1 (52) and CDKN2A/B (31). BCR-ABL1-like ALL has been identified in 10%–15% of B-ALL cases >1 year of age, but is more frequent in adolescents and young adults (28, 31, 45, 53, 54). Similarly, focal deletions of the oncogene ERG, now recognized as DUX4-rearranged, have been identified in 4%–7% of childhood B-ALL patients and 10% of adolescent B-ALL patients (20, 28, 55, 56).

Chromosomal microarrays, and other cytogenomic and molecular genetic techniques, have highlighted the occurrence of genetic deletions in the tumor suppressor gene, CDKN2A, the B-cell lineage regulator PAX5, and Ikaros family members IKZF1 and IKZF2 in B-ALL. All have been studied with respect to childhood ALL immunophenotype and cytogenomic subtype. Biallelic CDKN2A deletions have been reported in approximately 25% of B-ALL cases at the time of diagnosis with marked variation between cytogenomic subtypes (57). Hyperdiploid, MLL-rearranged, and ETV6-RUNX1 cases experience lower deletion rates (<15% each) while over 40% of cases with TCF3 (E2A)-Pbx1 or Ph+ (BCR-ABL1) harbor CDKN2A deletions (57). In T-ALL cases, biallelic CDKN2A deletions are more common than observed for B-ALL with approximately 50% of cases harboring a deletion and there is variation by subtype: 43% of TAL1-rearranged cases had a CDKN2A deletion as did 76% of cases with TLX3 rearrangements (57).

Deletions involving PAX5 have been observed in nearly one third of B-ALL cases with variation in frequency by ALL cytogenomic subtype (58, 59). Approximately 12% of ALL cases have deletions of IKZF1 (58, 60). Deletions are more common among B-ALL (13%) than T-ALL (5%) cases and variation by cytogenomic subtype has been reported with over 80% BCR-ABL1-positive (58–60) and 70%–80% of BCR-ABL1-like ALL (53) cases harboring an IKZF1 deletions. Mutations in Ras family members are also found in 20%–30% of childhood ALL cases (61). Mutations are found more frequently in Latino children and among children diagnosed with high hyperdiploid ALL (61). While characterizing deletion and mutation frequencies of the aforementioned genes in conjunction with cytogenomic abnormalities has begun, data on variation in deletion status by age, race, and sex are lacking.

Analytic Epidemiology of ALL

Because the hypothesized in utero origin of childhood leukemia, as well as prima facie evidence based on the detection of fusion genes including MLL-AF4, ETV6-RUNX1, and E2A-PBX1 in the neonatal blood spots of children diagnosed with ALL (40–42, 62, 63), many epidemiologic studies examine exposures from the preconception period through early life. Generally, gestational risk factors are thought to influence development of a “first hit” while early-life risk factors are thought to impart the “second hit” necessary for overt leukemic transformation. This may be especially relevant for the ETV6-RUNX1 subtype where it is believed that this fusion is necessary, but insufficient by itself, to lead to leukemia (64); whereas, MLL rearrangement alone appears to be sufficient to cause leukemia based on the identification of MLL rearrangements in blood spots of children diagnosed with ALL and the short latency between birth and diagnosis for a majority of cases with this subtype (42). The risk factor associations reported for ALL combined vary by window of exposure and by parent exposed. Figure 2 and Supplementary Table S1 display the measures of association for risk factors in association with ALL overall timing of the exposure during the preconception, pregnancy, birth, and postnatal periods. To identify articles for nongenetic risk factors, we gathered the meta- or pooled analyses measures of association for each exposure from the reports with the largest number of studies on each topic from PubMed (last accessed August 31, 2018). We used the search terms “acute lymphoblastic leukemia,” “meta-analysis,” and/or “Childhood International Leukemia Consortium.” Only the most recent meta-analyses and pooled analyses with three or more studies are displayed in Fig. 2 and Supplementary Table S1 (65–92).

The risk factors imparting an increased risk of childhood ALL occur for parental exposures to chemicals including pesticides [maternal residual pesticide exposure during pregnancy odds ratio (OR): 1.43; 95% confidence interval (CI): 1.32–1.54; refs. 66, 69] and paint (any parental residential paint exposure preconception OR: 1.54; 95% CI: 1.28–1.85; ref. 72) may impart the necessary “second hits” for some leukemic clones or lead to the accumulation of somatic mutations leading to disease initiation. Nonchemical exposures associated with an increased risk of ALL include maternal coffee intake during pregnancy of greater than or equal to 2 cups per day (OR: 1.27; 95% CI: 1.09–1.48; ref. 74), increasing maternal age (5-year increased OR: 1.05; 95% CI: 1.01–1.10; ref. 81), high birth weight (1 kg increase OR: 1.18; 95% CI: 1.12–1.32; ref. 82), large for gestational age at birth (OR: 1.21; 95% CI: 1.11–1.32; ref. 83), and prelabor cesarean delivery (OR: 1.23; 95% CI: 1.04–1.47; ref. 84). Conversely, postnatal exposures such as daycare attendance (any vs. none OR: 0.76; 95% CI: 0.67–0.87; ref. 91), being breastfed (any vs. none OR: 0.91; 95% CI: 0.84–0.98; refs. 88, 94), contact with dogs (any vs. none OR: 0.92; 95% CI: 0.86–0.99), and cats (any vs. none OR: 0.87; 95% CI: 0.80–0.94; ref. 90) in the first year of life, and the presence of allergies (any vs. none OR: 0.67; 95% CI: 0.54–0.82; ref. 86) have been found to reduce the risk of childhood ALL. The findings for daycare attendance and being breast fed are robust across multiple populations.
epidemiologic studies and are thought to provide a protective effect by priming the immune system early in life thereby allowing the individual to effectively control the development of leukemia through enhanced immune function as reviewed by Greaves (2018; ref. 95). The associations with maternal alcohol use (65) and smoking during pregnancy and ALL overall are null (80).

**Analytic Epidemiology of ALL by Subtype**

This section summarizes the literature on risk factors for ALL first by age at diagnosis and then by immunophenotype and cytogenomic subtype where available (Supplementary Table S1; Fig. 2). There may be differences in subtyping methods where CGH or FISH have evolved over time; therefore, between-study variation in subtyping may exist. Conversely, karyotyping methods, which have not changed dramatically over time, may be more consistent between studies. While studying differences in risk factor associations by subtype can identify etiologic variation among tumors, amassing enough childhood ALL cases, uniformly subtyping cases using contemporary definitions based on molecular profiling techniques, and obtaining detailed risk factor data have posed a tremendous challenge for the field as evidenced by the small number of studies with risk factor information for each clinically relevant subtype presented in Supplementary Table S1 and Fig. 2A–D. To identify articles for nongenetic risk factors stratified by immunophenotype and cytogenetic subtype, we first examined the meta- or pooled-analyses included, as described above, for each risk factor to determine whether risk estimates were given by ALL subtype. If no subtype-specific risk estimates were listed in the main text or available Supplementary Materials, we then examined the component studies included in each meta- or pooled-analysis to see whether they contained risk estimates stratified by subtype. If none of the component studies presented subtype-stratified risk estimates, we searched in PubMed (last accessed August 31, 2018) using the search terms "childhood acute lymphoblastic leukemia," "risk factors," and "subtype" to identify other studies where subtype-stratified risk estimates were presented either in the main text or Supplementary Materials. If more than one study reported subtype-stratified estimates, the risk estimates from the study with the most subtypes reported in their article or the study with the largest population size was presented.

**Risk of ALL by age at diagnosis**

Studies stratified by age at diagnosis were the first to shed light on the etiologic heterogeneity between subtypes of ALL because, as mentioned previously, age may serve as a proxy for some cytogenomic subtypes. For example, high birth weight was associated with an increased risk of infant ALL, which is dominated by the *MLL*-rearranged subtype, (OR = 2.8 for highest birth weight category in children <2 years of age; refs. 96, 97). In another study examining the trend in increasing birth weight by age at diagnosis, there were significant associations between each 1-kg increase and the risk of childhood ALL for children ages 1–5 years of age (OR: 1.24; 95% CI: 1.08–1.42) and 6–9 years of age (OR: 1.35; 95% CI: 1.05–1.74; ref. 98). This association was elevated, although not significant, for infants aged <1 years (OR: 1.62; 95% CI: 0.89–2.96), which may be due to a low sample size of low birth weight children <1 year of age, and was not associated with ALL among children ≥10 years of age at diagnosis (OR: 1.21; 95% CI: 0.90–1.63; ref. 98).

Paternal smoking prior to conception has shown variation in risk by age at ALL diagnosis in the offspring with strong effects observed for children diagnosed between 0 and 4 years of age (smokers vs. nonsmokers OR: 1.8; 95% CI: 1.2–2.6), null effects in children aged 5–9 years at diagnosis (OR: 0.9; 95% CI: 0.5–1.5), and 10–14 years at diagnosis (OR: 1.9; 95% CI: 0.5–1.8; ref. 99). It has also been observed that older maternal age and higher maternal education are associated with ALL in children aged 2–4 years (maternal age ≥35 vs. 20–34 years OR: 1.31, 95% CI: 1.06–1.63; maternal education ≥16 vs. 12–15 years OR: 1.48, 95% CI: 1.14–1.93) but not children aged 0–2 years at diagnosis (maternal age ≥35 vs. 20–34 years OR: 1.07, 95% CI: 0.71–1.62; maternal education ≥16 vs. 12–15 years OR: 1.01, 95% CI: 0.65–1.56; ref. 100). As hinted at in analyses examining differences in risk factors by age at diagnosis, there appears to be considerable etiologic heterogeneity in childhood ALL. This becomes even more apparent as we examine risk factor associations by the established immunophenotype and cytogenomic subtypes of ALL.

**Risk of ALL by immunophenotype**

Concerning B- and T-ALL, some studies report risk factor associations stratified by immunophenotype (Fig. 2A–D); however, while some risk factors are significantly associated with B-ALL, few risk factors show significant associations with T-ALL. The lack of significant risk factor associations by T-ALL may be due to the small sample size available for this minor subtype, which can be inferred from the increased width of the 95% CIs for T-ALL associations (sample sizes listed in Supplementary Table S1, where available). In addition, a number of risk factors are associated with ALL overall but are not associated with the ALL immunophenotypes, which makes it hard to discern whether there may be etiologic heterogeneity by ALL immunophenotype (73, 77, 81, 101–106). A number of risk factors display significant associations with B-ALL, but not T-ALL, as shown in Fig. 2A–D and Supplementary Table S1, including home paint exposure prior to conception, during pregnancy, and after birth (72), preconception paternal smoking (107), older paternal age (105), prelabor caesarean delivery (84), increasing birth weight (98, 108), increasing birth order (98), being breast fed (109), infections during childhood (89, 110), and daycare attendance (91, 111–113). A number of risk factors show similar associations with B-ALL and T-ALL (Fig. 2A–D; Supplementary Table S1). Exposures associated with significantly increased risks of both B- and T-ALL include large for gestational age (83), paternal occupational pesticide exposure (66), residential pesticide exposure during the preconception period, and during childhood (73). Exposures associated with significantly reduced risks of B- and T-ALL include any contact with cats (90) and prenatal vitamin use (114).

**Risk of ALL by cytogenomic subtype**

When cytogenomic subtype of B-ALL is considered, there is some variation in risk (Supplementary Table S1; Fig. 2A–D), but only a small number of studies report analyses stratified by these characteristics and consistency of the subtypes included is lacking across studies. Home paint exposure during pregnancy is associated with *ETV6-RUNX1* (OR: 1.51; 95% CI: 1.08–2.11) and any *MLL*-rearranged ALL (OR: 3.30; 95% CI: 1.71–6.35; ref. 72). Home paint exposure after birth is also associated with low hyperdiploid (OR: 1.94; 95% CI: 1.32–2.89) and *ETV6-RUNX1*.
B-ALL (OR: 1.60; 95% CI: 1.16–2.21; ref. 72). Daycare attendance is inversely associated with hyperdiploid B-ALL (OR: 0.38; 95% CI: 0.24–0.59; ref. 113) and ETV6-RUNX1 B-ALL (OR: 0.47; 95% CI: 0.24–0.94; ref. 113). Being breastfed is inversely associated with hyperdiploidy (OR: 0.37; 95% CI: 0.18–0.77; ref. 115). Increasing birth weight is significantly associated with hyperdiploid (OR: 1.12; 95% CI: 1.05–1.20) and TCF3 (E2A)-PTEN; t (1;19) ALL (OR: 1.41; 95% CI: 1.09–1.84; ref. 116). While maternal tea intake during pregnancy is not associated with an increased risk of ALL overall or by either immunophenotype, it was found to be associated with any MLL-rearranged B-ALL (≥2 cups/day vs. none OR: 2.69, 95% CI: 1.17–6.21); however, this association is based on a small number of cases (n = 9) born to mothers that consumed ≥2 cups/day and should be interpreted cautiously (74).

Risk of ALL by molecularly defined mutation subtype

Maternal smoking is one of the most commonly studied risk factors for childhood ALL to date and has also been investigated in a molecular ALL subtype analysis, apparently the first study of its kind, in relation to any exposure. In a case-only study, that assessed the association between tobacco exposure during the early-life period and deletions in 8 genes commonly lost somatically in childhood ALL (CDKN2A, ETV6, IKZF1, PAX5, RB1, BTG1, PAR1 region, and EBF1), a positive association was observed between maternal smoking and the total number of deletions observed in these genes in tumor samples from cases of childhood ALL (117). Interestingly, the authors observed that B-ALL and hyperdiploid ALL experienced fewer deletions of these genes while T-ALL and ETV6-RUNX1 B-ALL cases experienced one or more deletions in the genes under study. These findings suggest that maternal smoking may exert subtype-specific effects on ALL development but require validation in a case–control study. In a different analysis on parental exposures to solvents, plastic materials, and hydrocarbons, maternal exposure to these chemicals during pregnancy was associated with K-ras mutations among childhood ALL cases (118). These studies highlight the possibility of the simultaneous evaluation of cytogenomic and molecular subtypes in association with epidemiologic risk factors.

Germline Variants Display Subtype-Specific Variation in Association with ALL

Investigations of germline risk for ALL have generally involved genome-wide association studies (GWAS), which examine common variants, and more recently, whole-exome or genome-sequencing studies, which examine rare variants. Currently, GWAS comprise the bulk of germline genetic research on ALL (37, 119–124). The GWAS studies have been conducted among French (37), European (119, 120, 122), Australian (121), and American [white, black, and Hispanic (123); and Latino and non-Latino (124)] children. In addition, numerous replication studies have been done whereby stratification by immunophenotype and cytogenomic subtype has been performed, including some analyses in non-white populations. We present single-nucleotide variants (SNV) identified in discovery GWAS analyses, reported in replication analyses where stratification by subtype was performed, in association with childhood children only, as there was little racial variation observed in risk for most of the reported SNVs (Supplementary Table S2), and where cell counts were available to allow for the calculation of the per variant allele odds ratios (OR) and 95% confidence intervals (95% CI; Fig. 3). Supplementary Table S2 contains per allele ORs for SNVs in association with ALL overall and by subtype. GWAS studies were identified in PubMed (last accessed August 31, 2018) using the search terms “acute lymphoblastic leukemia” and “genome wide association study” and from references lists of current publications. Validation studies of individual SNVs were identified by searching for the specific gene along with the terms “acute lymphoblastic leukemia” and ‘subtype’ in PubMed.

ARID5B, AT-Rich Interaction Domain 5B, SNVs were identified in multiple GWAS studies with effect estimates (ORs) ranging from 1.5 to 1.9 per risk allele (37, 119–121, 124). ARID5B, which is involved in differentiation of B- ALL progenitors (119), SNV rs10821936 is associated with a significantly increased risk of ALL overall (OR: 1.9; 95% CI: 1.6–2.2; ref. 119) and this finding is consistent across race groups and sexes (24, 123, 125–133). This association persists among B-ALL only cases (OR: 1.83; 95% CI: 1.42–2.36; refs. 127, 128, 129, 132, 134, 135). The association is null between ARID5B SNV rs10821936 and T-ALL (OR: 0.96; 95% CI: 0.60–1.34) among Caucasian children (129, 132), but shows an increased association among black children (OR: 2.80; 95% CI: 1.16–6.77; ref. 24). Strong associations between rs10821936 and hyperdiploid B-ALL (OR: 2.94; 95% CI: 2.04–4.24; refs. 24, 127, 129, 132) and MLL germline (OR: 2.80; 95% CI: 1.60–5.00), but not MLL-rearranged (OR: 0.90; 95% CI: 0.60–1.40) cases have been reported (133). The IKAROS family zinc finger 1, IKZF1, gene, which is a transcription factor involved in lymphoid cell differentiation, particularly of T cells (136, 137), confers risk of ALL in pediatric cases (SNV rs4132601 OR: 1.7; 95% CI: 1.6–1.8; refs. 37, 120). When stratified by subtype, rs4132601 is associated with an increased risk of B-ALL regardless of race/ethnicity (Caucasian OR: 1.70; 95% CI: 1.46–1.99; refs. 132, 134, 138). Another gene harboring germline variants found to increase the risk of childhood ALL is CEBPE, which has a role in the determination of myeloid cells (139). GWAS-identified CEBPE SNV rs2239633 is associated with an increased risk of ALL overall (OR: 1.3; 95% CI: 1.2–1.5; refs. 120, 124) and among Caucasian B-ALL cases only (OR: 1.45; 95% CI: 1.25–1.68; ref. 138). In analyses where subtype of ALL was considered, associations with rs2239633 and disease status are largely null among infants and young children with and without MLL rearrangements (133, 140).

GATA3 SNV rs3824662 was studied in association with ALL overall (124) and by numerous cytogenomic subtypes (141, 142). Although GATA3 plays a role in T-cell differentiation (143), inherited germline variation in rs3824662 was significantly associated with B-ALL (OR: 1.31; 95% CI: 1.21–1.41) and hyperdiploid ALL (OR: 1.17; 95% CI: 1.06–1.29; ref. 141), although the strongest association was observed for Ph-like ALL (OR: 3.14; 95% CI: 1.18–5.44; ref. 142). Finally, SNVs within the following genes: PP4K2A (rs10828317), LHPP (rs35837782), EK3 (rs4762284), and CDKN2A (rs3731217) have been identified in GWAS studies as increasing the risk of childhood ALL (122, 124); however, none of these exhibit notable variation in association when considered by ALL overall (122, 124) relative to associations by immunophenotype or cytogenomic subtype (122, 144).

Studies examining rare germline variants have assessed mutations in TP53 in association with childhood ALL (145) and PAX5 in association with familial ALL (146). In a study of pathogenic TP53 variants among childhood ALL cases, 2% of cases had pathogenic TP53 mutations and 65% of pediatric patients...
Challenges in Identifying Etiologic Heterogeneity in ALL

Tremendous progress has been made in the treatment of and subsequently the survival from childhood ALL over the past few decades; however, less progress has been made in identifying modifiable risk factors, which may shed light on the initiation of leukemogenesis or aid in prevention. Meta- and pooled analyses, which allow for increased sample sizes, have been vital in identifying risk factors for childhood ALL overall and have identified pesticide exposure, paternal smoking, paint exposure, solvent exposure, and maternal coffee intake during pregnancy as potentially modifiable risk factors. On the other hand, immune-stimulating exposures such as being breastfed, attending daycare, having contact with cats and dogs, and the presence of allergies display inverse associations in childhood ALL. In studies where stratification by subtype of ALL has been presented, there is suggested variation in association by subtype; however, studies with larger sample sizes to accommodate the less frequent subtypes, such as T-ALL, are necessary to uncover any true etiologic heterogeneity by ALL subtype. Furthermore, the paucity of epidemiologic risk factor studies examining variation in risk by subtype highlights the need for future studies to examine etiologic heterogeneity using modern subtype definitions to keep pace with the clinically recognized subgroups. Despite the cytogenomic and molecular subtype definitions being clinically useful and prognostically different, studying risk factor differences between cytogenomic subgroups of ALL remains a challenge due to sample size constraints.

Figure 3.
Allelic odds ratios (OR) and 95% confidence intervals (95% CI) for single nucleotide variants (SNV) in association with childhood ALL overall and by subtype from discovery GWAS and replication analyses (reference numbers in brackets). Estimates reported for Caucasians only. Non-HD refers to nonhyperdiploid ALL.
identify risk-reduction strategies. The challenge of applying molecular subtyping methods to childhood ALL, as has been done in adult tumors, such as breast cancer, lies in the rarity of the disease and the rarity of each subtype. Therefore, a comprehensive childhood ALL study that has an adequate sample size to allow for stratification by the clinically relevant subtypes discussed herein and that also has performed uniform cytogenetic subtyping will be necessary to study etiologic heterogeneity by ALL subtype. To overcome the issues of small sample size and to achieve uniform subtyping, we recommend a nationwide recruitment strategy through clinical trial organizations such as the Children’s Oncology Group in the United States and the International BFM study group in Europe [147]. In addition, it is important to develop a means to investigate immune contribution to ALL by way of group in Europe (147). In addition, it is important to develop a means to investigate immune contribution to ALL by way of

subtyping, we recommend a nationwide recruitment strategy over clinical trial organizations such as the Children’s Oncology Group in the United States and the International BFM study

children ALL study that has an adequate sample size to allow for

focus on thoroughly studying the etiologic heterogeneity of childhood ALL by immunophenotype and cytogenetic subtype, which has been slower than the progress observed in adult tumor types.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

This work was supported by the NIH (grant number T32 CA099936; to L.A. Williams).

The costs of publication of this article were delayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 18, 2018, revised September 19, 2018; accepted February 5, 2019, published first February 15, 2019.

References


www.aacajournals.org Cancer Epidemiol Biomarkers Prev; 28(5) May 2019 853

Downloaded from cebp.aacrjournals.org on April 24, 2021. © 2019 American Association for Cancer Research.
Cancer Epidemiol Biomarkers Prev; 28(5) May 2019

Cancer Epidemiology, Biomarkers & Prevention

44. Wartenberg D, Groves FD, Adelman AS. Hematologic malignancies: acute
49. Strehl S, K
50. Chiaretti S, Ford AM, Le Beau MM,  ____1____
39. Raimondi SC, Roberson PK, Pui CH, Behm FG, Rivera GK. Hyperdiploid


39. Graves MF, Wiemels J. Origins of chromosome translocations in child-


46. Multilhagen CG, Miller CB, Radke J, Philips LA, Dalton J, Ma J, et al. BCR-


Is There Etiologic Heterogeneity between Subtypes of Childhood Acute Lymphoblastic Leukemia? A Review of Variation in Risk by Subtype

Lindsay A. Williams, Jun J. Yang, Betsy A. Hirsch, et al.


Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-18-0801

Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2019/02/15/1055-9965.EPI-18-0801.DC1

This article cites 144 articles, 28 of which you can access for free at:
http://cebp.aacrjournals.org/content/28/5/846.full#ref-list-1

This article has been cited by 2 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/28/5/846.full#related-urls

Sign up to receive free email-alerts related to this article or journal.

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

To request permission to re-use all or part of this article, use this link
http://cebp.aacrjournals.org/content/28/5/846.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.