Inherited genetic variation in childhood acute lymphoblastic leukemia

Takaya Moriyama,1,2 Mary V. Relling,2 and Jun J. Yang2

1Department of Pediatrics, Mie University Graduate School of Medicine, Mie, Japan; and 2Department of Pharmaceutical Sciences, St. Jude Children’s Research Hospital, Memphis, TN

Although somatically acquired genomic alterations have long been recognized as the hallmarks of acute lymphoblastic leukemia (ALL), the last decade has shown that inherited genetic variations (germline) are important determinants of interpatient variability in ALL susceptibility, drug response, and toxicities of ALL therapy. In particular, unbiased genome-wide association studies have identified germline variants strongly associated with the predisposition to ALL in children, providing novel insight into the mechanisms of leukemogenesis and evidence for complex interactions between inherited and acquired genetic variations in ALL. Similar genome-wide approaches have also discovered novel germline genetic risk factors that independently influence ALL prognosis and those that strongly modify host susceptibility to adverse effects of antileukemic agents (eg, vincristine, asparaginase, glucocorticoids). There are examples of germline genomic associations that warrant routine clinical use in the treatment of childhood ALL (eg, TPMT and mercaptopurine dosing), but most have not reached this level of actionability. Future studies are needed to integrate both somatic and germline variants to predict risk of relapse and host toxicities, with the eventual goal of implementing genetics-driven precision-medicine approaches in ALL treatment. (Blood. 2015;125(26):3988-3995)

Introduction

Acute lymphoblastic leukemia (ALL) is the most common cancer in children, accounting for 25% of all childhood malignancies.1,2 Sentinel chromosomal abnormalities (translocations or aneuploidy) are characteristic of the majority of ALL cases, and recent genomic profiling of leukemic cells continues to broaden our appreciation of the complex genomic landscape of this disease.3,6 These somatically acquired genomic aberrations are unique to ALL tumor cells; however, patients also carry inherited genetic variations (ie, germline variants) that are present in both normal and tumor cells. Although somatic genomic alterations have long been recognized as the hallmarks of ALL subtype classification, the last decade has shown that germline genetic variations are important determinants of interpatient variability in ALL susceptibility, drug response, and toxicities of ALL therapy (Table 1).

Common types of inherited genetic variations include single nucleotide polymorphisms (SNPs), insertions and deletions (gain or loss of short segments of sequence, indels), and structural variations (gain or loss of large segments of sequence; eg, copy number changes). Practically, nonmalignant cells from patients (eg, peripheral blood cells obtained during clinical remission) generally serve as the primary source of “germline” DNA. Recent advances in high-throughput genotyping technology enable agnostic screens of genetic variation across the entire human genome, with up to a few million genetic markers tested per patient. These “genome-wide” association studies, often referred to as GWASs do not rely on prior knowledge to focus on any subset of genes but, instead, systematically examine genetic variations in an unbiased fashion for their association with the phenotype of interest.7

Because of the large number of variants tested in GWASs, the required level of significance for association between a variant and a phenotype is generally set very high (P < 5 × 10−7) rather than the typical level of .05 for most power calculations.8,9 Thus, it is not surprising that there is limited power to detect genotype-phenotype associations using genome-wide approaches, and only genomic variation with great impact (large effect sizes) can be expected in most ALL GWASs. For example, with a sample size of 1000 patients, at an α level of 5 × 10−7, a frequency of the genomic variant in the population of 10%, and a phenotype that occurs in just 5% of patients (eg, such as central nervous system relapse of ALL or ALL therapy-related pancreatitis), in order to have 80% power to detect the genotype-phenotype association, the genomic variant would need to confer 4.5-fold higher risk of the trait than the wild-type or “normal” allele. For phenotypes or alleles that are less common, effect sizes would have to be even higher than 4.5-fold (or the sample size would need to be greater). Thus, in discovery studies, every effort must be made to minimize variation in nongenetic risk factors and to maximize sample size to improve the chance of observing associations between genomic variation and the phenotype of interest.

It should also be noted that commercial genotyping platforms that have been used in GWASs predominantly focus on relatively common genomic variants to achieve an even representation across all chromosomes, although with varying degrees of coverage and resolution.10 Most of these variants are intronic and may not be directly functional; instead, they are in at least partial linkage with other variants that are likely biologically active.11 As a result, findings from GWASs often require extensive follow-up studies to discover the true causal genetic variants underlying the GWAS signal. Although SNPs are the primary focus of GWASs, copy number variations can also be detected by most genome-wide SNP chip/arrays12 (except for small indels; eg, promoter repeats in TTMS).
Inherited genetic basis of ALL susceptibility

The risk of developing ALL is highest between 2 and 5 years after birth, with initiating sentinel somatic genomic lesions (eg, translocations) detectable at the time of birth in many cases. This early disease onset suggests a strong inherited genetic basis for ALL susceptibility. Inherited genetic risk factors for cancer can be divided into 2 main classes: rare penetrant variants associated with a high risk (may be observed in families with multiple members affected by ALL) and common less-penetrant variants associated with a modestly increased risk of ALL (such as those observed in population studies of ALL risk).

Rare germline mutations and familial ALL

A number of inherited genetic variants have been identified in excess in rare cases of familial ALL. For example, 50% of children with low-hypodiploid ALL have germline TP53 mutations characteristic of Li-Fraumeni cancer syndrome, an autosomal dominant familial

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP ID</th>
<th>Study design</th>
<th>Phenotype</th>
<th>Sample size (N)</th>
<th>Odds ratio (95% CI)</th>
<th>P</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL susceptibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARID5B</td>
<td>rs7089424</td>
<td>GWAS</td>
<td>ALL risk</td>
<td>907</td>
<td>1.65 (1.54-1.76)</td>
<td>6.7 × 10^-13</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>rs10821936</td>
<td>GWAS</td>
<td>ALL risk</td>
<td>441</td>
<td>1.91 (1.62-2.2)</td>
<td>1.4 × 10^-15</td>
<td>28</td>
</tr>
<tr>
<td>IKZF1</td>
<td>rs4132601</td>
<td>GWAS</td>
<td>ALL risk</td>
<td>907</td>
<td>1.69 (1.58-1.81)</td>
<td>1.2 × 10^-19</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>rs11978267</td>
<td>GWAS</td>
<td>ALL risk</td>
<td>441</td>
<td>1.69 (1.41-1.9)</td>
<td>8.8 × 10^-11</td>
<td>28</td>
</tr>
<tr>
<td>CEBP2</td>
<td>rs2239633</td>
<td>GWAS</td>
<td>ALL risk</td>
<td>907</td>
<td>1.34 (1.22-1.45)</td>
<td>2.9 × 10^-7</td>
<td>27</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>rs17756311</td>
<td>GWAS</td>
<td>ALL risk</td>
<td>2450</td>
<td>1.36 (1.18-1.56)</td>
<td>1.4 × 10^-5</td>
<td>30</td>
</tr>
<tr>
<td>PIP4K2A</td>
<td>rs7088318</td>
<td>GWAS</td>
<td>ALL risk</td>
<td>2450</td>
<td>1.40 (1.28-1.53)</td>
<td>1.1 × 10^-11</td>
<td>30</td>
</tr>
<tr>
<td>GATA3</td>
<td>rs3824662</td>
<td>GWAS</td>
<td>ALL risk</td>
<td>3107</td>
<td>1.31 (1.21-1.41)</td>
<td>8.6 × 10^-12</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWAS</td>
<td>Risk for Ph-like ALL</td>
<td>511</td>
<td>3.85 (2.7-5.4)</td>
<td>2.2 × 10^-14</td>
<td>32</td>
</tr>
<tr>
<td>TP63</td>
<td>rs17505102</td>
<td>GWAS</td>
<td>Risk for ETV6-Runx1 ALL</td>
<td>1370</td>
<td>0.65 (0.52-0.75)</td>
<td>8.9 × 10^-9</td>
<td>39</td>
</tr>
<tr>
<td>Treatment outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPMT</td>
<td>rs1800462</td>
<td>Candidate gene</td>
<td>Minimal residual disease</td>
<td>814</td>
<td>0.34 (0.13-0.86)</td>
<td>.02</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>rs1800460</td>
<td>Candidate gene</td>
<td>Minimal residual disease</td>
<td>814</td>
<td>0.34 (0.13-0.86)</td>
<td>.02</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>rs1142345</td>
<td>Candidate gene</td>
<td>Minimal residual disease</td>
<td>601</td>
<td>0.36 (0.15-0.88)</td>
<td>.03</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>rs1800460</td>
<td>Candidate gene</td>
<td>Relapse</td>
<td>487</td>
<td>2.67 (1.53-4.68)</td>
<td>8.9 × 10^-7</td>
<td>55</td>
</tr>
<tr>
<td>IL15</td>
<td>rs17007695</td>
<td>GWAS</td>
<td>Minimal residual disease</td>
<td>2535</td>
<td>3.61 (2.34-5.57)</td>
<td>6.7 × 10^-9</td>
<td>58</td>
</tr>
<tr>
<td>PYGL</td>
<td>rs7142143</td>
<td>GWAS</td>
<td>Relapse</td>
<td>2535</td>
<td>1.41 (1.22-1.64)</td>
<td>5.1 × 10^-6</td>
<td>58</td>
</tr>
<tr>
<td>PDE4B</td>
<td>rs663977</td>
<td>GWAS</td>
<td>Relapse</td>
<td>781</td>
<td>1.43 (1.10-1.86)</td>
<td>.007</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>rs3824662</td>
<td>GWAS</td>
<td>Relapse</td>
<td>710</td>
<td>1.38 (1.03-1.83)</td>
<td>.039</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GWAS</td>
<td>Relapse</td>
<td>2258</td>
<td>2.0 (1.71-3.66)</td>
<td>2.3 × 10^-6</td>
<td>31</td>
</tr>
<tr>
<td>Toxicities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPMT</td>
<td>rs1800462</td>
<td>Candidate gene</td>
<td>Thiopurine-induced</td>
<td>180</td>
<td>9.3 (3.58-24.27)</td>
<td>.007</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>rs1800460</td>
<td>Candidate gene</td>
<td>myelosuppression</td>
<td>180</td>
<td>9.3 (3.58-24.27)</td>
<td>.007</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>rs1142345</td>
<td>Candidate gene</td>
<td>myelosuppression</td>
<td>180</td>
<td>9.3 (3.58-24.27)</td>
<td>.007</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>rs16685232</td>
<td>GWAS</td>
<td>Thiopurine intolerance</td>
<td>657</td>
<td>*</td>
<td>8.8 × 10^-9</td>
<td>70</td>
</tr>
<tr>
<td>ACP1</td>
<td>rs12714403</td>
<td>GWAS</td>
<td>Glucocorticoid-induced osteonecrosis</td>
<td>362</td>
<td>5.6 (2.7-11.3)</td>
<td>1.9 × 10^-6</td>
<td>80</td>
</tr>
<tr>
<td>GRIA1</td>
<td>rs4958351</td>
<td>GWAS</td>
<td>Asparaginase allergy</td>
<td>485</td>
<td>1.75 (1.41-2.17)</td>
<td>3.5 × 10^-7</td>
<td>84</td>
</tr>
<tr>
<td>HLA-DRB1</td>
<td>HLA-DRB1*07:01</td>
<td>Candidate gene</td>
<td>Asparaginase allergy</td>
<td>1870</td>
<td>1.64 (1.28-2.09)</td>
<td>7.5 × 10^-5</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>rs116855232</td>
<td>GWAS</td>
<td>Asparaginase allergy</td>
<td>502</td>
<td>2.92 (1.82-4.80)</td>
<td>1.4 × 10^-5</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>rs3832526</td>
<td>Candidate gene</td>
<td>Asparaginase allergy</td>
<td>533</td>
<td>14.6 (3.6-58.7)</td>
<td>&lt;.0005</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>rs1056892</td>
<td>Candidate gene</td>
<td>Asparaginase pancreatitis</td>
<td>8.6 (2.0-37.3)</td>
<td>.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs1056892</td>
<td>Candidate gene</td>
<td>Anthracycline-induced</td>
<td>487</td>
<td>1.79 (1.08-2.96)</td>
<td>.02</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>rs2322228</td>
<td>GWAS</td>
<td>Anthracycline-induced</td>
<td>362</td>
<td>3.7 (1.3-10.2)</td>
<td>.05</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>rs924607</td>
<td>GWAS</td>
<td>Anthracycline-induced</td>
<td>321</td>
<td>*</td>
<td>4.7 × 10^-8</td>
<td>90</td>
</tr>
<tr>
<td>SLCO1B1</td>
<td>rs11045879</td>
<td>GWAS</td>
<td>Methotrexate clearance</td>
<td>640</td>
<td>*</td>
<td>8.2 × 10^-11</td>
<td>104, 105</td>
</tr>
<tr>
<td></td>
<td>rs4149056</td>
<td>Candidate gene</td>
<td>Methotrexate clearance</td>
<td>206</td>
<td>16.4 (8.7-26.7)</td>
<td>.004</td>
<td>104, 105</td>
</tr>
<tr>
<td></td>
<td>rs4149056</td>
<td>Candidate gene</td>
<td>Methotrexate clearance</td>
<td>115</td>
<td>*</td>
<td>.008</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>rs4149056</td>
<td>Candidate gene</td>
<td>Methotrexate clearance</td>
<td>415</td>
<td>*</td>
<td>3.5 × 10^-4</td>
<td>50</td>
</tr>
</tbody>
</table>

CI, confidence interval; GI, gastrointestinal; GWAS, genome-wide association study; ID, identification; Ph, Philadelphia chromosome; SNP, single-nucleotide polymorphism.
cancer syndrome characterized by a range of other solid and brain tumors. Germline mutations of PAX5, which encodes a transcriptional factor required for B-cell differentiation, were also found in 2 unrelated kindreds, each of which had 5 family members develop ALL.16 However, the vast majority of childhood ALL is not familial, and TP53 or PAX5 mutations represent a very small population attributable risk (ie, proportion of ALL cases that can be explained by these risk factors).

Common variants and susceptibility to childhood ALL

Common genetic variants influencing leukemia susceptibility can be identified by association studies comparing the frequency of variations in unrelated ALL cases vs controls (individuals not affected by ALL); variants overrepresented in cases may contribute to the risk of developing this disease (examples given in Table 1). There is an extensive body of work that examines the contribution of a number of “candidate” pathways (eg, carcinogen metabolism, folate metabolism, DNA repair) to ALL risk, but with oftentimes conflicting results. A recent meta-analysis summarized 47 studies of 25 polymorphisms in 16 genes and observed statistically significant \( P < 0.05 \), albeit modest, associations with ALL susceptibility for 8 variants (eg, CYPIA1*2A and XRCC1 G28152A).17 However, it should be noted that the false-positive probability in this study was estimated at 20%. Similar pooled analyses subsequently confirmed the association for multiple variants in CYPIA1 and XRCC1,18,19 although with some variability by ancestry and age. Several epidemiology studies noted significant associations between infection and risk of ALL in children, pointing to potential roles of host immune defense in ALL etiology.20-22 In fact, germline SNPs at the HLA-DF and HLA-DOA loci were associated with ALL susceptibility in admixed populations in the United States.23,24 However, a comprehensive analysis of the major histocompatibility complex region in 824 B-ALL cases and 4737 controls of European genetic ancestry did not find statistically significant association signals in this genomic region after correcting for multiple testing.25 Caution needs to be exercised when examining HLA variants, especially in diverse populations, because of the complex linkage disequilibrium and excessive diversity at these loci in different races and ethnic groups. Variants in IL15, IL12A, and other genes related to adaptive immunity were also reported to potentially predispose children to ALL, although further validation is warranted.26,27

The first pair of GWASs of childhood ALL susceptibility were published in 2009, independently identifying ARID5B, IKZF1,27,28 and CEBPE27 as genome-wide significant risk loci in children of European descent. Subsequent GWASs with larger sample sizes and/or greater population diversity discovered additional susceptibility variants in CDKN2A, BM11-PIP4K2A, and GATA3.29-32 Unlike candidate-gene studies, these GWAS hits have been repeatedly validated by subsequent reports.33-40 Interestingly, genomic loci implicated by ALL susceptibility GWASs are often also targeted by somatic genomic aberrations in ALL cells. For example, IKZF1, an important transcription factor in all lymphoid lineages, is frequently deleted in ALL blast cells (particularly in high-risk ALL), which confers a poor prognosis.4 Loss of CDKN2A/CDKN2B tumor suppressor genes also occurs in up to 40% of B-precursor ALL and contributes to cell cycle deregulation in leukemia.5 However, there does not appear to be any cosegregation of germline ALL risk variants and somatic abnormalities involving the same gene, suggesting that inherited and acquired variations occur and function independently.

Of 6 genome-wide significant ALL risk loci, lead variants in ARID5B, IKZF1, GATA3, and PIP4K2A are significant regardless of genetic ancestry, whereas the effects of CEBPE and CDKN2A variants were more restricted to Europeans.30,32,38 Also, frequencies of ALL risk alleles at ARID5B, PIP4K2A, and GATA3 differ significantly by ancestry in a pattern that is consistent with racial differences in ALL incidence (Africans < Europeans < Hispanics), and are therefore likely to contribute to ancestry-related differences in ALL susceptibility.

ALL consists of subgroups with different genomic abnormalities, each of which may have distinct genetic susceptibility. Initial GWASs have already noted considerable differences in the effects of susceptibility variants by ALL molecular subtype. For example, an ARID5B variant is significantly overrepresented in ALL cases with hyperdiploid karyotypes and less so in children with T-cell ALL.28,38 A PIP4K2A variant was also enriched in hyperdiploid ALL among B-cell ALL.30,31,37 In populations of European descent, variants in the TP63 gene were genome-wide significantly associated with the acquisition of the t(12;21) translocation in ALL.39 Similarly, intronic variants in GATA3 strongly influence the risk of developing Ph-like ALL and were also associated with the risk of relapse.35 A contemporaneous GWAS also identified these GATA3 SNPs overrepresented in childhood ALL cases with high-risk clinical features (older age and higher leukocyte count at diagnosis), although the Ph-like phenotype was not explicitly ascertained in this study.41 These data collectively illustrate the complex interactions between genetic variations in the host (inherited) and those in ALL cells (acquired) and their unique contributions to disease pathogenesis and treatment outcomes.

It is fair to argue that these GWASs have produced unequivocal evidence for an inherited genetic basis of ALL susceptibility. However, the molecular mechanisms by which these variants are linked to ALL risk are largely unknown. For example, the vast majority of susceptibility variants identified are intronic, and their effects on gene functions are not clearly understood. In some instances (eg, rs3824662 in the GATA3 gene), the risk variant is located in a genomic region rife with enhancer elements active in hematopoietic tissues and is directly linked to GATA3 transcription. Therefore, we posit that ALL risk loci identified by GWASs are likely to overlap with regulatory DNA elements in the genome, possibly influencing gene function by modulating transcription. Future functional studies are needed to describe the details of these molecular processes.

Germline genetic variation and ALL treatment outcome

Although the survival rates of childhood ALL increased significantly in the past few decades due to risk-directed therapy, there is still substantial variation in treatment response, with 15% of children with ALL experiencing relapse.41 In fact, ALL relapse is the fifth most common cancer in children and a leading cause of death in this cancer. The interindividual variation in relapse risk can arise from both tumor- and host-related factors. Gene expression profiling and, more recently, whole-genome sequencing studies discovered tumor genetic features associated with outcome and drug resistance.6,15,42,43 In parallel, there is increasing evidence that inherited genetic variations play important roles in determining patients’ risk of relapse (Table 1).

Candidate genes related to response to ALL therapy

Inherited genetic variation can contribute to ALL treatment response by influencing host disposition of antileukemic agents, interactions between ALL and tumor, and tumor biology itself. In particular, it was widely hypothesized that variation in genes involved in antileukemic
drug metabolism would be associated with treatment outcome of ALL therapy. For example, patients with loss-of-function variants in the TPMT gene had significantly lower levels of minimal residual disease (MRD) compared with those with wild-type TPMT after 2 weeks of therapy including mercaptopurine.44 In a subsequent study of 601 children treated on the Nordic Society of Paediatric Haematology and Oncology ALL-92 protocol, TPMT deficiency was associated with a lower risk of relapse, plausibly due to higher levels of active mercaptopurine metabolites in patients with defective TPMT.55 In contrast, TPMT genotype was not predictive of hematologic relapse risk in the St Jude Children’s Research Hospital (St Jude) Total Therapy XIIIB protocol, most likely because mercaptopurine dose was already individualized on the basis of TPMT status to achieve comparable exposure to active metabolites.46 More recently, a 2.9-kb intronic germline deletion in the BIM gene was shown to alter the splicing pattern and consequently result in the loss of proapoptotic isoforms of BIM, required for glucocorticoid cytotoxicity in ALL.47,48 This intronic deletion of BIM in ALL cells also conferred significant resistance to dexamethasone,49 although the exact impact of this polymorphism on ALL relapse risk in patients remained unclear. Other candidate-gene studies have identified relapse risk variants in MTHFR, TYMS, GSTM1, and ABCC4, but the degree of association at these loci varied significantly among studies, plausibly due to differences in ALL treatment regimens.50-54

GWASs of ALL treatment outcome

In 2009, Yang et al reported one of the first GWASs of ALL treatment response in which the authors identified 102 SNPs associated with end-of-induction MRD in 487 children with newly diagnosed ALL on St Jude and Children’s Oncology Group (COG) frontline clinical trials.55 Twenty percent of the MRD-related SNPs were also associated with pharmacokinetics and pharmacodynamics of antileukemic agents, generally linking the same allele to MRD eradication and greater drug exposure. In particular, germline intronic variants in IL15 were consistently associated with MRD in both cohorts; these SNPs positively regulate IL15 expression, and higher IL15 levels protect hematologic cancer cells from cytotoxic agents.56 A recent independent report confirmed that autocrine and paracrine IL15 signaling led to significant growth advantage of primary B-precursor ALL cells in vitro through induction of STAT3, ERK1/2, and to a lesser extent PI3K and NF-κB signaling.57 A subsequent GWAS focused on relapse risk in 2535 children with ALL and discovered 134 relapse-related SNPs, of which 133 (99%) remained prognostic after adjusting for known relapse risk factors (ALL subtypes defined by tumor cytogenetics, age, and leukocyte count at diagnosis, and MRD).58 The top-ranked hit in this study was an intronic variant in the PYGL gene, which was associated with a 3.6-fold higher risk of relapse (P = 6.7 × 10⁻⁹). Glycogen phosphorylase (PYGL) is a target of adenosine monophosphate, which plays a critical role in response to antileukemic agents such as mercaptopurine and methotrexate. Also notable was the highly significant association with relapse observed for PDE4B variants. Prior studies have already shown that inhibition of PDE4B induces apoptosis in chronic lymphoblastic leukemia and diffuse large cell lymphoma59,60 and sensitizes cells to glucocorticoid-induced cell death.51,62 In ALL, pharmacologic inhibition of PDE4 results in growth suppression and dexamethasone sensitivity.63 suggesting glucocorticoid response as a plausible mechanism by which PDE4B is linked to ALL relapse. In a more recent study of 34,000 preselected potentially clinically relevant SNPs in 778 European children with newly diagnosed ALL, the authors discovered 11 cross-validated SNPs associated with relapse risk.64 Combined analyses of host genomic profiles, clinical presenting features, and MRD status further identified 3 distinct risk groups with highly divergent prognoses.

Germline genetic variants characteristic of Native American ancestry have been associated with increased risk of ALL relapse, explaining the inferior treatment outcome in children with ALL of self-declared Hispanic ethnicity.65 Ancestry-related poor prognosis was abrogated by the addition of a single extra phase of chemotherapy (delayed intensification), pointing to the potential utility of treatment individualization based on germline genetic variants. In fact, the aforementioned susceptibility variants in GATA3 for Ph-like ALL are significantly overrepresented in individuals with higher Native American genetic ancestry (characteristic of self-reported Hispanics), potentially contributing to ancestry-related differences in ALL relapse.53 These GATA3 variants were associated with MRD and relapse in 2 cohorts of children treated on COG frontline protocols, which was also true in >2000 children enrolled on the Berlin-Frankfurt-Munster clinical trials for newly diagnosed ALL.31

Taken together, both candidate-gene and genome-wide studies have identified inherited genetic variations related to interpatient variability in ALL treatment outcomes. However, the extent to which the effects of inherited germline variants on MRD and relapse are confounded by (or independent of) ALL tumor genetic factors is unclear, and integrated analyses including both germline and somatic genetic variations will hopefully provide comprehensive characterization of genetic risk factors for ALL relapse.

Pharmacogenomics of adverse effects of ALL therapy

Discovering the genomic basis for adverse effect phenotypes in ALL is complicated by the fact that all drug-induced phenotypes will be at least partly dependent on drug therapy; thus, it is critical to control for variability in drug exposure when conducting studies to elucidate the genomic basis of the adverse effect. Because relatively subtle differences among ALL regimens can have substantial impacts on the frequency and severity of adverse effects and because most ALL regimens differ from each other (eg, drugs used, doses used, combinations, and schedules), the power to detect genomic influences on adverse-effect phenotypes is diminished as each treatment group is added as a stratification variable, in that effective sample size decreases with each new grouping. Other covariates that must be included in analyses of how genotype variation may influence adverse-effect phenotypes include genomic ancestry and, often, age. Because collection of germline DNA has not been a routine component of many ALL trials, and not all ALL trials routinely capture adverse-effect phenotypes, the field is still in its infancy in terms of discovering genetic variants that are associated with ALL adverse effects.

Although some adverse effects (eg, myelosuppression) due to ALL therapy can be linked to a number of antileukemic agents, some can largely be linked to specific drugs. These include glucocorticoid-induced osteonecrosis, vincristine neuropathy, anthracycline cardiomyopathy, asparaginase-induced allergy and pancreatitis, and methotrexate-induced mucositis and neurotoxicity. There have been candidate-gene and genome-wide approaches to identify inherited variants that can explain some of the risk of these drug-specific adverse effects in ALL (Table 1). Interestingly, although myelosuppression can be caused by many agents, a substantial portion of myelosuppression during continuation therapy is due to a monogenic defect in TPMT,66,68 which has led to the use of TPMT genetic testing
to modify starting dosages of thiopurines.\textsuperscript{69} More recently, a coding variant in \textit{NUDT15} has been reported to account for thiopurine intolerance, particularly in those with East Asian ancestry and of Hispanic ethnicity,\textsuperscript{70,71} arguing that contemporary ALL treatment regimens (eg, drug dose) developed in populations of European descent may require modifications to be appropriate for non-European populations due to differences in genetic variations.

**Glucocorticoids**

Osteonecrosis is associated with glucocorticoid use. It has been hypothesized that several mechanisms can lead to the loss of blood supply to bone, which causes the ultimate phenotype, including in some cases thrombosis and hyperlipidemia. It is likely that additional treatment-related factors (eg, asparaginase)\textsuperscript{72-73} play a role not only in the incidence but also in the mechanism of glucocorticoid-related osteonecrosis, and given the strong association with adolescent age, it is possible that some genetic risk factors may be more penetrant in some age groups than in others. Candidate-gene studies have implicated inherited variation in \textit{HAS3} and \textit{VDR} genes related not only to osteonecrosis but also to hypoalbuminemia and hypercholesterolemia (supportive of a role for drug-induced lipodystrophy) as contributors to osteonecrosis risk.\textsuperscript{80} Additional genome-wide studies for osteonecrosis risk in the setting of differing age groups and differing ALL therapeutic protocols are needed to define genetics of this disorder.

**Asparaginase**

Asparaginase use has increased in several recent ALL regimens, bolstered by data indicating that relapses are prevented by increased asparaginase exposure.\textsuperscript{81-83} Although its frequency has decreased with the more common use of pegylated formulations, up to 40% of patients develop allergy to asparaginase. Asparaginase allergy is detrimental not only because of morbidity associated with allergy, but because allergy is associated with lower serum asparaginase concentrations and because asparaginase doses may be missed and thus therapy can be compromised. In a frontline St Jude trial,\textsuperscript{84} the top-ranked SNP associated with allergy was in \textit{GRIAI} on chromosome 5q33. SNPs in this locus have previously been associated with asthma and atopy in non-ALL settings.\textsuperscript{85} In a larger study of St Jude and COG patients, \textit{HLA} variants were imputed using genome-wide SNP data and external reference sets; the \textit{HLA-B:07:01} variant was associated with asparaginase allergy and the presence of antibodies against asparaginase, and the variants were predicted to alter binding between \textit{HLA} proteins and asparaginase epitopes.\textsuperscript{86} Using a candidate-gene approach and pooling together the reactions of allergies, pancreatitis, and thrombotic events, it has been reported that variants in \textit{ASNS} were associated with these asparaginase-related adverse effects.\textsuperscript{87}

**Anthracyclines**

The risk of cardiomyopathy from anthracyclines has been assessed in long-term survivors including those treated for ALL. Candidate-gene studies implicated \textit{CBR3} in the risk of cardiomyopathy, particularly at lower doses of anthracyclines;\textsuperscript{88} patients exposed to higher doses were at high risk of cardiomyopathy, regardless of genotype. Broader genomic studies, using a platform directed at cardiovascular variants, identified that \textit{HAS3} predisposed to cardiomyopathy, most strongly in those exposed to higher anthracycline doses.\textsuperscript{89} These findings illustrate the principle that pharmacogenetic risk factors may be highly dependent on the exact therapeutic regimen, with some genetic risk factors most evident at lower drug doses and others most evident at higher drug doses.

**Vincristine**

Vincristine neuropathy can be a major dose-limiting adverse effect in ALL. In a genome-wide study, a higher frequency of neuropathy has been associated with a promoter variant in \textit{CEP72} (rs924607).\textsuperscript{90} The frequency of the risk allele was lower in individuals with African ancestry compared with the other ancestral groups, consistent with a lower incidence of vincristine neuropathy in African American patients.\textsuperscript{91} A candidate-gene study found that variants in \textit{ABCB1}, \textit{ACTG1}, and \textit{CAPG} were associated with vincristine neurotoxicity during ALL therapy,\textsuperscript{92} although other candidate-gene studies found no associations with \textit{ABCB1} variants, despite its likely role in vincristine transport.\textsuperscript{93,94} Although \textit{CYP3A5} affects vincristine metabolism, candidate-gene studies indicate that there are conflicting data on its association with neuropathy.\textsuperscript{92,93,95,96}

**Methotrexate**

There have been extensive pharmacogenetic studies of methotrexate in ALL.\textsuperscript{97-99} Candidate-gene studies have focused on common variants in genes clearly involved in the folate pathway, such as \textit{MTTHFR}, \textit{SLC19A1}, \textit{TYMS}, and \textit{DHFR} but,\textsuperscript{90,99} despite multiple candidate-gene studies for toxicity, results have been conflicting (or based on single, nonreplicated small studies), and thus it is currently not possible to recommend changes to methotrexate dosing based on inherited variants in these candidate genes.\textsuperscript{97,98} Genome-wide studies identified variants associated with leukoencephalopathy,\textsuperscript{100} but these findings have not yet been replicated. Methotrexate effects are influenced by interindividual variation in its plasma clearance, leading some to implement an approach that targets systemic exposure based on clearance.\textsuperscript{101-103} Genome-wide analyses identified multiple common genomic variants in \textit{SLCO1B1} that were associated with methotrexate clearance,\textsuperscript{104} a finding that has been replicated in several studies\textsuperscript{100,106} and confirmed in preclinical models.\textsuperscript{107,108} The high degree of replication for \textit{SLCO1B1} variants as a determinant of methotrexate clearance stands in contrast to the lack of replicated findings using a candidate-gene approach.\textsuperscript{97,98}

**Perspectives**

Studies of germline genomic determinants in ALL have multiple objectives, one of which is to gain new biological insights into the mechanisms of leukemogenesis or ALL response (desired antileukemic effects or host toxicities) that could eventually yield improvements in diagnosis or therapy. Another, more elusive objective, is to discover genetic variation that can itself be used as a diagnostic or therapeutic test. For example, it is possible that tests of germline \textit{TP53} status can be used in families of patients with hypodiploid ALL to provide risk estimates for individuals in the family. Likewise, germline tests of \textit{TPMT} status can be used for individualizing the dose of thiopurines to minimize host toxicity without adversely affecting outcomes.\textsuperscript{69} Currently, there are relatively few germline genomic associations that have the required level of evidence on clinical utility to permit routine use as a clinical test. However, the field is likely to change as new data emerge over the next few years, especially with the rapid advances in next-generation sequencing.
that the exciting possibility of exhaustively interrogating all variants in the genome (eg, rare variants with large effects).

Ultimately, one can foresee that somatically acquired ALL-specific genetic alterations as well as inherited genomic variants will be used to predict each patient’s risk of relapse and host toxicities with differing treatment regimens, and the choice of treatment protocol can be informed by balancing the probability of cure vs the probability of adverse effects based on genetic and other patient characteristics. For example, patients carrying highly penetrant germline variants related to life-threatening toxicities (pancreatitis) may be considered for treatment regimens that are not highly dependent on asparaginase, especially if his/her germline and/or tumor genetic profiles indicate sensitivity to other chemotherapeutics. Conversely, optimizing antileukemic effect is weighted more in patients with high-risk ALL, particularly if they are predicted to experience modest toxicities based on germline genetic variations. The delicate balance between toxicity and efficacy in this context is challenging, and large collaborations are needed to comprehensively evaluate outcome- or toxicity-related genetic variants in diverse treatment regimens and to develop genetics-based decision support systems. Childhood ALL is uniquely positioned for this type of translational research, given the impressive progress already made in genomics and pharmacogenomics of this disease and the exceptionally organized clinical trials for children with ALL.

Acknowledgments

This work was supported by the National Institutes of Health National Institute of General Medical Sciences grant U01 GM092666 and National Cancer Institute grants CA176063, CA142665, CA21765, CA36401, and CA156449; Leukemia & Lymphoma Society grant 6168-12; and the American Lebanese Syrian Associated Charities. J.J.Y. is an American Society of Hematology Scholar. T.M. is supported by the Study-Abroad Scholarship of Mie Prefecture, Japan.

Authorship

Contribution: J.J.Y., M.V.R., and T.M. conceived of and wrote the manuscript.

Conflict-of-interest disclosure: M.V.R. receives royalties from licensing TPMT genotyping (Prometheus Laboratories). The remaining authors declare no competing financial interests.

Correspondence: Jun J. Yang, Pharmaceutical Sciences MS313, St. Jude Children’s Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105-3678; e-mail: jun.yang@stjude.org.

References


Inherited genetic variation in childhood acute lymphoblastic leukemia

Takaya Moriyama, Mary V. Relling and Jun J. Yang