Genome-wide Interrogation of Germline Genetic Variation Associated With Treatment Response in Childhood Acute Lymphoblastic Leukemia

Jun J. Yang, PhD
Cheng Cheng, PhD
Wenjian Yang, PhD
Deqing Pei, MS
Xueyuan Cao, MS
Yiping Fan, PhD
Stanley B. Pounds, PhD
Geoffrey Neale, PhD
Lisa R. Tremain, PhD
Deborah French, PhD
Dario Campana, MD, PhD
James R. Downing, MD
William E. Evans, PharmD
Ching-Hon Pui, MD
Meenakshi Devidas, PhD
W. P. Bowman, MD
Bruce M. Camitta, MD
Cheryl L. Willman, MD
Stella M. Davies, MBBS, PhD
Michael J. Borowitz, MD, PhD
William L. Carroll, MD
Stephen P. Hunger, MD
Mary V. Relling, PharmD

Context  Pediatric acute lymphoblastic leukemia (ALL) is the prototype for a drug-responsive malignancy. Although cure rates exceed 80%, considerable unexplained interindividual variability exists in treatment response.

Objectives  To assess the contribution of inherited genetic variation to therapy response and to identify germline single-nucleotide polymorphisms (SNPs) associated with risk of minimal residual disease (MRD) after remission induction chemotherapy.

Design, Setting, and Patients  Genome-wide interrogation of 476,796 germline SNPs to identify genotypes that were associated with MRD in 2 independent cohorts of children with newly diagnosed ALL: 318 patients in St Jude Total Therapy protocols XIIIB and XV and 169 patients in Children’s Oncology Group trial P9906. Patients were enrolled between 1994 and 2006 and last follow-up was in 2006.

Main Outcome Measures  Minimal residual disease at the end of induction therapy, measured by flow cytometry.

Results  There were 102 SNPs associated with MRD in both cohorts (median odds ratio, 2.18; \( P = .0125 \)), including 5 SNPs in the interleukin 15 (IL15) gene. Of these 102 SNPs, 21 were also associated with hematologic relapse (\( P < .05 \)). Of 102 SNPs, 21 were also associated with antileukemic drug disposition, generally linking MRD eradication with greater drug exposure. In total, 63 of 102 SNPs were associated with early response, relapse, or drug disposition.

Conclusion  Host genetic variations are associated with treatment response for childhood ALL, with polymorphisms related to leukemia cell biology and host drug disposition associated with lower risk of residual disease.

JAMA. 2009;301(4):393-403 www.jama.com

©2009 American Medical Association. All rights reserved.

PEDIATRIC ACUTE LYMPHOBLASTIC leukemia (ALL) cure rates have increased from less than 10% in the 1960s to more than 80% today. Such advancement was partly derived from the identification of presenting clinical features (eg, molecular subtype, leukocyte count, age) associated with treatment outcome and subsequent implementation of risk-adapted therapy.\(^1,2\) The assessment of decreasing disease burden in response to therapy by sequential monitoring of minimal residual disease (MRD) status has now been integrated into risk stratification.\(^3,5\) Minimal residual dis-
ease assays provide a direct assessment of early treatment response and are associated with final treatment outcome.6-9 Response to treatment varies during the 4- to 6-week phase of remission induction therapy, as exemplified by changes in early sequential MRD assays.4,8,9 Thus, some patients exhibit drastic depletion of leukemia (from 100% to less than 0.01% leukemia cells in bone marrow) within only 2 to 3 weeks of induction therapy, while others exhibit high levels of residual leukemia even after 4 to 6 weeks of therapy.

This interindividual variation in treatment response in cancer can arise from both tumor- and host-related factors; however, most prior studies focused on the former. Gene expression profiling of leukemic blasts has identified tumor genetic features associated with outcome10,11 and drug resistance.12-15 Much less is known about host genetic factors associated with cancer cure rates.16-19 Taking a global approach to identify host genetic factors that may affect treatment response in ALL, we tested germline single-nucleotide polymorphisms (SNPs) for their association with MRD at the end of remission induction therapy in 2 independent cohorts of children treated for newly diagnosed ALL.

### METHODS

#### Patients

Two cohorts of patients were included (Table 1) with approval of their respective institutional review boards. Written informed consent from patients or their guardians (as appropriate) for genomic research was included as part of the treatment protocols at St Jude (Total Therapy protocols XIIIB and XV) and the treatment/biology protocol for the COG (Pediatric Oncology Group 9900 study).

#### Treatment and MRD Assessment

There were common and unique elements to the eligibility and treatment of the St Jude and COG cohorts

**Table 1. Patient Characteristics and Relation to MRDa**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>St Jude Cohort</th>
<th></th>
<th>Children’s Oncology Group Cohort</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Patients</td>
<td>Patients in the MRD Genome-wide Association Study</td>
<td>P Value for Relation to MRDb</td>
<td>All Patients</td>
</tr>
<tr>
<td>Racec</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>282 (76)</td>
<td>246 (77)</td>
<td>.63</td>
<td>132 (58)</td>
</tr>
<tr>
<td>Black</td>
<td>58 (16)</td>
<td>45 (14)</td>
<td></td>
<td>12 (5)</td>
</tr>
<tr>
<td>Other</td>
<td>21 (6)</td>
<td>27 (9)</td>
<td></td>
<td>63 (27)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>213 (57)</td>
<td>181 (57)</td>
<td>.33</td>
<td>154 (68)</td>
</tr>
<tr>
<td>Female</td>
<td>158 (43)</td>
<td>137 (43)</td>
<td></td>
<td>73 (32)</td>
</tr>
<tr>
<td>Age at diagnosis, y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;1</td>
<td>6 (1)</td>
<td>2 (1)</td>
<td>.28</td>
<td>73 (32)</td>
</tr>
<tr>
<td>1-10</td>
<td>262 (71)</td>
<td>236 (74)</td>
<td></td>
<td>154 (68)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>103 (28)</td>
<td>80 (25)</td>
<td></td>
<td>103 (45)</td>
</tr>
<tr>
<td>White blood cell count at diagnosis, /µL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50,000</td>
<td>269 (73)</td>
<td>237 (75)</td>
<td>.07</td>
<td>124 (55)</td>
</tr>
<tr>
<td>≥50,000</td>
<td>102 (27)d</td>
<td>81 (25)</td>
<td></td>
<td>103 (45)</td>
</tr>
<tr>
<td>Lineaged</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B lineage</td>
<td>299 (81)</td>
<td>248 (78)</td>
<td>.06</td>
<td>227 (100)</td>
</tr>
<tr>
<td>T cell</td>
<td>72 (19)</td>
<td>70 (22)</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Molecular subtype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEL-AML1</td>
<td>65 (17)</td>
<td>61 (19)</td>
<td>.33</td>
<td>3 (1)</td>
</tr>
<tr>
<td>BCR-ABL</td>
<td>12 (3)d</td>
<td>0</td>
<td></td>
<td>23 (10)d</td>
</tr>
<tr>
<td>E2A-PBX1</td>
<td>13 (3)</td>
<td>0</td>
<td></td>
<td>18 (8)d</td>
</tr>
<tr>
<td>MLL rearrangement</td>
<td>7 (2)</td>
<td>0</td>
<td></td>
<td>183 (81)</td>
</tr>
</tbody>
</table>

**Abbreviation:** MRD, minimal residual disease.

*Data are presented as No. (%) of patients unless otherwise indicated.

*Association between patient characteristics and MRD assessed by χ² test (P<.05) in patients included in the genome-wide association study.

*Race was assigned based on germline genotype of approximately 600,000 SNPs as described in the Supplemental Methods (http://www.jama.com).

*Patient characteristics significantly associated with MRD by χ² test (P<.05).
signal intensity and consistency with expected genotypes based on linkage disequilibrium. SNPs with a minor allele frequency of less than 1% or call rates less than 95% (ie, the number of samples with a definitive genotype call at this SNP is <95% of the total number of samples typed for this SNP) were excluded (Figure 1); patient samples that failed to achieve 95% call rates (ie, samples for which <95% of interrogated SNPs were successfully typed) were excluded (Figure 1 and Supplemental Methods).

### Genome-wide Association Analysis for MRD

Minimal residual disease was treated as an ordinal variable; ie, 1 for negative, 2 for positive, and 3 for high-positive, as defined above. To minimize confounding effects, patients with ALL subtypes that were strongly related to MRD status and that differed in frequency between the 2 cohorts (ie, E2A-PBX1, MLL, or BCR-ABL) were excluded from the MRD analyses (Table 1, Figure 1, and Supplemental Methods). The final analysis included 476 796 SNPs among 318 St Jude and 169 COG patients (Table 1 and Figure 1).

SNPs associated with end-of-induction MRD were identified based on a bidirectional validation approach in both the St Jude and COG cohorts, comprising a 3-step analysis. Our goal was to find SNP genotypes that were associated with MRD in both cohorts—those that might be generalizable across induction treatment for ALL.

In step 1, we computed the statistical significance for each SNP genotype’s association with MRD in each cohort separately. The Spearman rank correlation was used as the test statistic to account for the ordinal nature of MRD and the gene dosage effect of genotypes. P values were computed by a permutation-asymptotic hybrid method (Supplemental Methods). An additive model was assumed, although the trend test is also reasonably robust to moderate deviation from additivity.

In step 2, we determined the threshold for statistical significance by estimation of the false discovery rate (FDR) and an internal validation (Supplemental Methods) in each cohort. Using the P values obtained in step 1, in each cohort, FDR levels were estimated on a grid of per-test significance levels (P value cutoffs). Based on the FDR estimates and the internal validation, a threshold (P ≤ .0125) was chosen for each cohort to declare a set of SNPs for further investigation.

In step 3, we used the COG MRD cohort to validate the top-ranked SNPs (P ≤ .0125) discovered in the St Jude MRD cohort and vice versa (bidirectional validation) using a rank-based inference procedure (Supplemental Methods). The 102 overlapping SNPs satisfying the significance threshold determined in step 2 (FDR estimation and internal validation) and step 3 (bidi-
rectional validation) were prioritized for further investigation and association with additional relevant phenotypes. The absolute risk difference was estimated for the 102 overlapping SNPs by comparing the frequency of MRD positivity in patients with 1 or 2 copies of the risk allele vs those homozygous for the alternative allele.

As a secondary approach, we used the St Jude cohort as a discovery set and the COG cohort as a test set (Supplemental Methods).

Operating characteristics of the Spearman rank correlation test were determined via a simulation study (Supplemental Methods and eFigure 4). The genotypes associated with MRD were also assessed by a pooled analysis that combined evidence across the 2 cohorts to provide a combined P value for each SNP (Supplemental Methods). The FDR, the false-positive report probability, and the population-attributable fraction (Supplemental Methods) for prioritized SNPs were estimated.

Statistical and computational analyses were performed using S-Plus software, version 7.0 (Insightful Corp, Seattle, Washington), R software, version 26.1 (http://www.r-project.org), and SAS software, version 9.1 (SAS Institute Inc, Cary, North Carolina). Data analyses were performed between 2006 and 2008.

Association of MRD SNPs With Additional Phenotypes

Antileukemic Response. The relationship between the 102 overlapping MRD SNP genotypes and 2 additional leukemia response phenotypes was analyzed to prioritize SNPs and to minimize the risk of false discoveries.

For purposes of this analysis, patients were also retrospectively categorized into super-responders, responders, and poor responders based on consideration of MRD status at 2 time points. Minimal residual disease status was dichotomized as negative or positive as defined above. Super-responders were MRD-negative at both early (day 8 in COG, day 19 in St Jude) and later (day 28 in COG, day 46 in St Jude) time points; responders were MRD-positive at the early time point but became MRD-negative at the later time point; and poor responders had positive status at the later time point. The association between SNP genotypes and this MRD responsiveness phenotype was assessed by rank correlation in all evaluable patients in separate analyses of St Jude (n=304) and COG (n=154).

The cumulative incidence of hematologic relapse (including isolated and combined hematologic plus extramedullary relapses) as a function of SNP genotypes in the combined St Jude and COG cohorts was analyzed by the Gray test. Isolated extramedullary relapses, lineage switch, second malignancy, and death in remission were treated as competing events. Excluding individuals with E2A-PBX1, MLL rearrangements, or BCR-PBX1 ALL, 416 St Jude and 180 COG patients were included in this analysis, overlapping with but not identical to the MRD cohorts as defined in Figure 1 and Table 1. Of these, 33 in St Jude and 35 in COG experienced hematologic relapse. St Jude patients were divided into 4 strata according to their treatment protocol and risk classification, and COG patients formed the fifth stratum. The cumulative incidence hazard regression model of Fine and Gray was used to confirm the directional association with relapse for SNPs that achieved P<.10 in the Gray test. Follow-up was confirmed as of March 2006 for St Jude patients and November 2006 for COG patients.

Pharmacokinetic Studies. Three pharmacokinetic phenotypes were available from a subset of St Jude patients for antileukemic agents used during remission induction. Patients in these 3 data sets overlapped with but were not identical to those studied in the primary St Jude cohort for MRD.

The first data set included plasma clearance of etoposide on day 29 of therapy in 97 patients enrolled in St Jude Total Therapy protocol XIIIB. Although etoposide was a component of induction therapy for only a subset of the St Jude MRD cohort and none of the COG cohort, its elimination is mediated via cytochrome P450 3A4 (CYP3A4) and P-glycoprotein, a common mechanism of elimination that also affects prednisone, vincristine, and anthracyclines.

The second data set included intravenous methotrexate plasma clearance at day 1 in 319 patients treated in St Jude Total Therapy protocols XIIIB and XV. The third data set included intracellular methotrexate polyglutamate accumulation in ALL blasts at 44 hours after receiving up-front methotrexate in 230 patients treated in St Jude trials. Although only a subset of the St Jude MRD cohort and none of the COG MRD cohort received intravenous methotrexate, all patients in both cohorts received intrathecal methotrexate, which is known to distribute from cerebrospinal fluid to blood systemically.

The relationship between SNP genotypes and pharmacokinetic variables was analyzed using linear regression.

RESULTS

Patients and MRD Status

From St Jude Total Therapy protocols XIIIB (accrual between 1994-1998) and XV (2000-2006), 371 children with newly diagnosed ALL had available germline DNA and evaluable MRD status at the end of induction therapy. Of the patients with ALL enrolled in the COG study P9906 (accrual between 2000 and 2003), 227 children had germline DNA and evaluable end-of-induction MRD status. The actual number of patients included in specific analyses is described below. We found no significant differences in the characteristics (age, initial leukocyte count, race, and sex) of the patients enrolled in the St Jude and COG trials who were not
was established based on FDR estimates and an internal validation inferene (Supplemental Methods and eFigure 2). Through a rank-based bidirectional validation, a significant impact of germline variation on MRD identified in the St Jude cohort was validated in the COG cohort ($P=2.2 \times 10^{-6}$) and that identified in the COG cohort was validated in the St Jude cohort ($P<10^{-11}$) (Supplemental Methods).

In total, 102 SNPs exhibited significant concordant association with end-of-induction MRD ($P \leq .0125$) in both the St Jude and COG cohorts, with a median odds ratio of 2.18 and a median population-attributable fraction of 0.17 (Table 2 for top 25 SNPs by combined cohort $P$ value; eTables 1 and 2 for full details). Among these 102 SNPs, 50 were annotated to genes. Because 45 SNPs were clustered at 15 genomic loci by linkage disequilibrium (pairwise $r^2>0.5$), these 102 SNPs represented 72 unique genomic loci (eFigure 3). A SNP in the ST8SIA6 (NM_001004470.1) gene (for combined cohort, odds ratio, 3.91; absolute risk difference: 0.46; $P=9.6 \times 10^{-5}$) had the strongest association with MRD but had no significant flanking SNPs and a relatively low minor allele frequency of 4% (Figure 2 [chromosome 1]). The next highest-ranked SNP (rs17007695) was in the interleukin 15 (IL15 (NM_172174.2) locus (Figure 2 [chromosome 4]; Table 2) and was notable for strong (for combined cohort, odds ratio, 2.67; absolute risk difference, 0.28;
GERMLINE GENETIC VARIATION AND TREATMENT RESPONSE IN CHILDHOOD LEUKEMIA

P = 8.8 \times 10^{-7}) and comparable association with MRD in both the St Jude (P = 4.4 \times 10^{-4}) and COG cohorts (P = 2.3 \times 10^{-4}). Moreover, this SNP was flanked by 4 IL15 SNPs (rs17015014, rs10519612, rs10519613, and rs35964658; Figure 3) that were also associated with MRD in both cohorts (Figure 3, Table 2, and eTable 1), and these 5 SNPs were in linkage disequilibrium with each other (pairwise r^2, 0.48-0.97).

Several of the highly ranked SNP genotypes have relatively small numbers of patients in the least common genotypic groups (eTable 2). Three (50%) of the 6 St Jude patients with the CC genotype, 35.6% of those with the CT genotype (n = 45), and only 15.8% of patients with the TT genotype (n = 267) at the IL15 SNP rs17007695 had detectable MRD at the end of induction therapy, with a similar finding observed in the COG cohort (Figure 4). The C allele at the IL15 germline SNP rs17007695 was weakly associated (P = .0701) with a higher IL15 expression in leukemic blasts, and overexpression of IL15 was associated with MRD in both cohorts (P = .0342 in St Jude and P = .0035 in COG; eTable 5).

All 102 SNPs remained significantly associated with MRD after adjustment for race, sex, leukocyte count at diagnosis, age, and ALL subtype (eTable 1). To further explore whether the genotypes had similar associations with MRD in each racial group, we examined the frequency of MRD positivity for each SNP genotype in each major racial group. An example is provided by the SNP rs13106616, illustrating that the GG genotype was similarly associated with a lower risk of MRD across 3 racial groups, although the allele frequency differed significantly by race (eTable 6). We also assessed the false-positive report probability for these 102 SNPs, and 82 (80.4%) exhibited a false-positive report probability of less than 0.5 (Table 1), a level associated with replicated associations in other contexts.26-33

Genome-wide Association Analysis for MRD Using the 2-Stage Discovery and Validation Approach

In addition to the bidirectional validation described herein, we also present a genome-wide analysis for SNPs associated with end-of-induction MRD by following the “discovery and validation” approach. In the discovery stage, we computed the statistical significance of each SNP genotype’s association with MRD in the discovery cohort (St Jude), estimating permutation-asymptotic hybrid P values for

---

**Figure 2. Overview of Genome-wide Association Results in the Combined St Jude and COG Analysis (n=487)**

**Figure 3. Association of IL15 SNPs With MRD**

Single nucleotide polymorphisms (SNPs) in the interleukin 15 (IL15) region plotted with their P values (shown as –log_{10}) from the correlation between genotype and minimal residual disease (MRD) combining the St Jude and Children’s Oncology Group (COG) minimal residual disease cohorts (details in the Supplemental Methods available at http://www.jama.com) and plotted from chromosomes 1 to X. Colors discriminate chromosomes.
association with MRD, as detailed in the Supplemental Methods. A P value threshold of \(7 \times 10^{-4}\) was arrived at by balancing the levels of false-negative and false-positive errors using the profile information criterion (eFigure 5)\(^2\); 624 SNPs met this threshold. In the second stage, these SNPs were then tested in the validation cohort (COG). Of these, 39 exhibited concordant associations at \(P \leq .05\), more than what would be expected by chance (\(P = .021\) by Fisher exact test), and those that overlap with the bidirectional approach are shown in Table 2 and eTable 1. A P value threshold of .0125 for the discovery cohort, 8635 SNPs met this cutoff, 330 of which were validated in the COG cohort with \(P \leq .05\), exceeding what would be expected by chance (\(P = 1.8 \times 10^{-9}\)).

### Relation of MRD-Associated SNPs to Other Antileukemic Response Phenotypes

Although end-of-induction therapy MRD is highly associated with long-term treatment outcome, an earlier reduction of leukemic burden is also informative.\(^3\) Thus, nearly all patients with negative MRD at early time points (day 19 in St Jude and day 8 in COG) remained leukemia-free. We examined which of the 102 overlapping SNPs could also distinguish patients who responded early (superresponders; \(n = 145\) in St Jude and \(n = 26\) in COG) vs those who remained MRD-positive at the end of induction therapy (poor responders; \(n = 59\) in St Jude and \(n = 52\) in COG) vs individuals who were MRD-positive at the early time point but MRD-negative later (responders; \(n = 100\) in St Jude and \(n = 76\) in COG). Of the 102 overlapping SNPs, 40 (\(40\%\)) were also associated (\(P < .05\)) with early response in both cohorts (eTable 3).

Of the 102 SNPs, 21 were significantly associated with hematologic relapse by stratified Gray test and in a cumulative incidence hazard regression model (\(P < .05\); eTable 3). An example is shown for rs1486649 (an intergenic SNP); there was a monotonic relationship between the number of copies of the C allele and the risk of hematologic relapse (Figure 5).

### Relation of MRD-Associated SNPs to Antileukemic Drug Pharmacokinetics

To understand mechanisms by which host genetic variation might affect treatment response, we tested whether the 102 overlapping SNP genotypes were related to antileukemic drug disposition in a set of St Jude patients evaluable for pharmacokinetics (Table 3; eTable 3). In total, 21 of the 102 MRD-related SNPs exhibited significant association with antileukemic agent pharmacokinetics, with 3 SNPs associated with more than 1 pharmacokinetic phenotype. Eight of 102 SNPs were associated with methotrexate clearance (at \(P < .05\)); all 8 genotypes were associated with positive MRD and greater drug clearance. Ten of the 102 SNPs were associated with the etoposide pharmacokinetics, with 7 of 10 associated with positive MRD and greater drug clearance.

Similarly, 6 of the 102 SNPs were significantly associated with the leukemic cell accumulation of methotrexate polyglutamates, with 5 of 6 associated with positive MRD and lower methotrexate polyglutamates. Thus, of 24 significant associations,
One of the strongest signals from the genome-wide scan came from 5 SNPs located in the IL15 locus, a proliferation-stimulatory cytokine. Interleukin 15 can protect hematologic tumors from glucocorticoid-induced apoptosis in vitro, and IL15 expression in ALL blasts has been linked to risk and relapse of CNS leukemia. Both higher IL15 expression (\(P = .0342\) in St Jude and \(P = .003\) in COG) and germline SNP genotypes were associated with an increased risk of positive MRD (eTable 5), and we found a trend (\(P = .0701\)) toward a relationship between IL15 SNP genotypes and IL15 expression in ALL leukemic blasts. Several of the IL15 SNPs that were associated with MRD have been linked to enhanced IL15 transcription/translation efficiency in vitro. Thus, it is plausible that germline genetic variation in IL15 plays a role in treatment response in childhood ALL via affecting IL15’s function or quantity in ALL blasts, and the fact that IL15 SNPs were prominent from unbiased genome scans in 2 independently treated cohorts points to its importance in ALL response, either as a prognostic marker or as a therapeutic target.

Because genome-wide interrogations for pharmacogenetics are still in their infancy, there are no published whole-genome data linking polymorphisms with anticancer drug response. We had the opportunity to couple the findings from our genome-wide SNP interrogation for MRD with 3 relevant host pharmacokinetic phenotypes: systemic clearance of 2 antileukemic agents (etoposide and methotrexate) and intracellular disposition of the latter. Although 4 to 8 different antileukemic agents were used in these 2 cohorts, remarkably, 21 of the 102 MRD-associated SNPs were also significantly associated with disposition of these 2 antileukemic agents. Although many additional genetic variations would be expected to be specific for antileukemic drugs other than methotrexate and etoposide and might therefore account for some of the remaining 81 MRD-associated SNPs, several of the pathways involved in methotrexate disposition and etoposide disposition (http://www.pharmgkb.org) are likely to be shared by other antileukemic agents. Particularly for etoposide, whose disposition involves CYP3A4 metabolism and P-glycoprotein excretion, it is likely that there is overlap in the genetic determinants of its disposition with those affecting anthracyclines, glucocorticoids, and vincristine. The majority (83.3%) of the associations

### Table 3. Examples of Relationships Between MRD-Associated SNPs and Host Disposition of Antileukemic Drugs

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Host Disposition of Drugs</th>
<th>No. of MRD-Positive Patients</th>
<th>No. of MRD-Negative Patients</th>
<th>MRD-Positive, %</th>
<th>(P) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP_A-2155892 (rs7992226)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>125 (103.1-143.4)</td>
<td>2</td>
<td>17</td>
<td>10.5</td>
<td>.01</td>
</tr>
<tr>
<td>GA</td>
<td>119 (92.4-150.6)</td>
<td>15</td>
<td>105</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>131 (104.4-166.8)</td>
<td>44</td>
<td>134</td>
<td>24.7</td>
<td></td>
</tr>
<tr>
<td>SNP_A-4236270 (rs9871556)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TT</td>
<td>2222 (890.0-4282.0)</td>
<td>10</td>
<td>66</td>
<td>13.2</td>
<td>.03</td>
</tr>
<tr>
<td>CT</td>
<td>2293 (1144.0-4474.0)</td>
<td>19</td>
<td>123</td>
<td>13.4</td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>1524 (609.1-2817.0)</td>
<td>32</td>
<td>68</td>
<td>32.0</td>
<td></td>
</tr>
<tr>
<td>SNP_A-2172039 (rs13106616)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>46.7 (41.01-53.14)</td>
<td>38</td>
<td>199</td>
<td>16.0</td>
<td>.003</td>
</tr>
<tr>
<td>GA</td>
<td>48.3 (45.44-60.0)</td>
<td>16</td>
<td>51</td>
<td>23.9</td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>56.2 (50.5-63.7)</td>
<td>7</td>
<td>7</td>
<td>50.0</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: MRD, minimal residual disease; SNP, single-nucleotide polymorphism.

aFor SNP_A-2155892, host disposition of antileukemic drugs shown for median methotrexate plasma clearance in mL/min/m\(^2\); for SNP_A-4236270, as median methotrexate polyglutamate accumulation in leukemic blasts in pmol/10\(^6\) cells; and for SNP_A-2172039, as median etoposide plasma clearance in mL/min/m\(^2\).
b\(P\) values indicate the statistical significance of the association between SNP genotype and pharmacokinetic phenotypes as determined by linear regression.

©2009 American Medical Association. All rights reserved.
between SNP genotypes and drug disposition were pharmacologically intuitive, with genotypes that were associated with increased drug exposure linked to lower levels of MRD. Together, these results suggest that more attention should be given to details of drug administration and risk factors for rapid drug clearance, in addition to the attention already placed on more granular risk classification of ALL.

There was also a high proportion (21/102) of SNPs that were associated with not only MRD but also with the risk of hematologic relapse in both cohorts. This high percentage is somewhat surprising in that the postremission therapy (which would ultimately be expected to have a significant effect on relapse risk) differed substantially in the COG and St Jude cohorts. This secondary analysis does lend credence to the hypothesis that we did identify true associations between SNP genotypes and poor response.

Like all risk features, genotypes that are informative for pharmacogenetic phenotypes are likely to be highly dependent on therapy. For this reason, we purposefully chose 2 cohorts (St Jude and COG) that had received somewhat different remission induction regimens, with slightly different time points for the primary phenotype (MRD), to identify polymorphisms more likely to have prognostic significance across multiple therapeutic regimens. Compared with the traditional “discovery and validation” approach, this bidirectional approach minimizes bias against the discovery cohort (more stringent P value cutoff). The disadvantage of this approach is that we might have missed SNPs more specific to the few elements of therapy that differed between the cohorts, and may have a higher FDR.

It is important to consider race, both from the standpoint of its possible effects on antileukemic drug efficacy and from its influence on germline SNP allele frequency. We found good agreement between self-declared race and that determined using ancestry-informative SNPs, and the 102 MRD-associated SNPs remained significant after adjusting for ancestry (eTable 1). The fact that SNP genotypes maintained significance after adjusting for race, despite substantial differences in some allele frequencies by race, suggests that inherent differences in ALL prognosis among racial groups are partly influenced by differences in allele frequencies among racial groups, which could in the future lead to “race-neutral” (but genomically based) individualization of therapy.

We acknowledge that despite the fact that these SNP genotypes were associated with MRD in 2 independent cohorts, there is a danger of false-negative and false-positive findings, especially when sample size is relatively small. However, phenotypes of interest in pharmacogenetic studies (eg, CYP2C9/VKORC1 for warfarin and TPMT for thiopurine) may have effect sizes that exceed those likely to be observed for multigenic common diseases, and, therefore, smaller sample sizes may suffice in the former. By identifying 102 SNPs based on association in 2 independent cohorts and by further validation of 62% of these SNPs (eTable 3) to be associated with the related phenotypes of relapse, “super response” at days 8 or 19, and antileukemic drug pharmacokinetics, we have further decreased the chance of false discoveries. The SNPs we identified may be in linkage disequilibrium with the truly causative genetic variants that have not yet been directly genotyped (eTable 4). Importantly, few of the 102 polymorphisms we identified have previously been suggested as candidates for affecting antancer drug efficacy, and approximately half of the genomic variants are not annotated to genes at all, illustrating the need to further explore mechanisms by which germline genomic variation affects interindividual variability in antileukemic drug response.

Although the acquired genetic characteristics of tumor cells play a critical role in drug responsiveness, our results show that inherited genetic variation of the patient also affects effectiveness of anticancer therapy, and that genome-wide approaches can identify novel and yet plausible pharmacogenetic variation. Such variation may be factored into treatment decisions in the future by placing additional emphasis on optimizing drug delivery to overcome host genetic variation, in addition to the current emphasis on tumor genetic variation.

Author Contributions: Dr Relling had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: J. Yang, Cheng, Evans, Camitta, Willman, Carroll, Relling.

Acquisition of data: W. Yang, Pei, Neale, French, Campana, Downing, Evans, Pui, Devidas, Bowman, Willman, Borowitz, Hunger, Relling.

Analysis and interpretation of data: J. Yang, Cheng, Pei, Cao, Fan, Pounds, Treviño, Evans, Camitta, Willman, Davies, Carroll, Hunger, Relling.

Drafting of the manuscript: J. Yang, Evans, Camitta, Willman, Relling.

Critical revision of the manuscript for important intellectual content: J. Yang, Cheng, W. Yang, Pei, Cao, Fan, Pounds, Neale, Treviño, French, Campana, Downing, Evans, Pui, Devidas, Bowman, Willman, Davies, Borowitz, Carroll, Hunger, Relling.

Statistical analysis: J. Yang, Cheng, W. Yang, Pei, Cao, Fan, Pounds, Devidas.

Obtained funding: Campana, Evans, Davies, Borowitz, Carroll, Relling.

Administrative, technical, or material support: Neale, Treviño, French, Campana, Downing, Evans, Pui, Carroll, Hunger, Relling.

Study supervision: Hunger, Relling.

Financial Disclosures: Dr Cheng serves on the Data Safety Monitoring Board for Enzon Pharmaceuticals, for which he is compensated. St Jude Children’s Research Hospital allocates a portion of the income it receives from licensing inventions and tangible research materials to those researchers responsible for creating this intellectual property. Under this policy, Drs Evans and Relling report that they receive a portion of the income St Jude receives from licensing patent rights related to TPMT polymorphisms and GGH polymorphisms. Dr Pui has received honoraria for lectures and scientific consultation from Sanofi-Aventis. Dr Borowitz reports that he has been a consultant for and receives support in kind from BD Biosciences and research support from Genzyme. Dr Relling reports that she is on the Scientific Advisory Board for Genome Explorations (uncompensated) and that she will likely soon receive funding for investigator-initiated research on the pharmacology of aspiraginase from Enzon Pharmaceuticals. No other financial disclosures were reported.

Funding/Support: This work was supported by grants CA 093592-02, CA 51001, CA 78224, CA 21765, CA R3736401, CA 80419, CA 806011, CA 98543, CA 29139, and National Institutes of Health (NIH)/National Institute of General Medical Sci-

©2009 American Medical Association. All rights reserved.
REFERENCES

45. Bleyer WA, Nelson JA, Kamen BA. Accumula-


