Published in final edited form as:
JAMA. 2009 January 28; 301(4): 393–403. doi:10.1001/jama.2009.7.

Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia

Jun Yang, PhD[1], **Cheng Cheng, PhD**[1], **Wenjian Yang, PhD**[1], **Deqing Pei, MS**[1], **Xueyuan Cao, MS**[1], **Yiping Fan, PhD**[1], **Stan Pounds, PhD**[1], **Lisa R. Treviño, PhD**[1], **Deborah French, PhD**[1], **Dario Campana, MD, PhD**[1], **James R. Downing, MD**[1], **William E. Evans, Pharm D**[1], **Ching-Hon Pui, MD**[1], **Meenakshi Devidas, PhD**[2], **W.P. Bowman, MD**[3], **Bruce M. Camitta, MD**[4], **Cheryl Willman, MD**[5], **Stella M. Davies, MBBS, PhD**[6], **Michael J. Borowitz, MD, PhD**[7], **William L. Carroll, MD**[8], **Stephen P. Hunger, MD**[9], and **Mary V. Relling, Pharm D**[1]

1*St. Jude Children's Research Hosp., Memphis, TN*

2*University of Florida, Gainesville, FL*

3*Cook Children's Medical Center, Ft. Worth, TX*

4*Medical College of Wisconsin, Milwaukee, WI*

5*University of New Mexico Cancer Center, Albuquerque, NM*

6*Cincinnati Children's Hospital and Medical Center, Cincinnati, OH*

7*Johns Hopkins Medical Institute, Baltimore, MD*

8*New York University Medical Center, NY, NY*

9*The Children's Hospital and the University of Colorado Cancer Center, Aurora, CO*

Abstract

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.

Objective—Using a genome-wide approach, 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—We performed a genome-wide interrogation of 476,796 germline SNPs to identify genotypes that predicted MRD in two independent cohorts of children with newly diagnosed ALL: 318 patients on St. Jude trials Total XIIIB and XV and 169 patients on a Children's Oncology Group (COG) trial P9906.

Main Outcome Measures—MRD at the end of induction therapy, measured by flow cytometry.

Correspondence: Mary V. Relling, Department Pharmaceutical Sciences, St. Jude Children's Research Hospital, 262 Danny Thomas Place MS313, Memphis, TN 38105 Phone: 901-495-2348, mary.relling@stjude.org.

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*: Yang J, Cheng, Relling; *Acquisition of data*: Campana, Downing, French, Devidas, Pui, Bowman, Willman, Borowitz, Relling; *Drafting of the manuscript*: Yang J and Relling; Critical revision of the manuscript for important intellectual content: Yang J, Cheng, Yang W, Pounds, Trevino, French, Evans, Pui, Camitta, Borowitz, Carroll, Hunger, Relling; *Statistical analysis*: Cheng, Yang W, Pei, Cao, Fan, Pounds; *Obtaining funding*: Campana, Evans, Borowitz, Carroll, Davies, Relling; *Study supervision*: Relling.

Results—There were 102 SNPs associated with MRD in both cohorts (P≤0.0125), including 5 SNPs in the interleukin 15 (*IL15*) gene. A high proportion, 21 of these 102 SNPs, also predicted hematologic relapse (P<0.05). 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 also associated with early response, relapse, or with drug disposition.

Conclusions—Host genetic variability affected treatment response for childhood ALL, and germline variants may exert their effects on MRD by effects on leukemic cell biology and on host disposition of antileukemic drugs.

Introduction

The past three decades have witnessed steady improvements in treatment of pediatric acute lymphoblastic leukemia (ALL), with cure rates increasing from less than 10% in the 1960s to over 80% today. Such drastic advancement was partly derived from the identification of presenting clinical features (e.g. molecular subtype, leukocyte count, age) predictive of 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] MRD assays provide a direct assessment of early treatment response and are predictive of final treatment outcome, even after adjusting for other prognostic factors.[6-9]

Response to treatment varies during the 4-6 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 their leukemia cells (from 100% to less than 0.01% leukemia cells in the bone marrow) within only 2-3 weeks of induction therapy, while others exhibit high levels of residual leukemia even after 4-6 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 diagnostic leukemic blasts has identified tumor cell genetic features associated with outcome[10,11] and drug resistance in childhood ALL.[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 affect treatment response in ALL, we interrogated 476,796 germline single nucleotide polymorphisms (SNPs) for their association with MRD at the end of remission induction therapy. We studied two independent cohorts: children with newly diagnosed ALL treated on protocols at St. Jude Children's Research Hospital (St. Jude) and through the Children's Oncology Group (COG). We discovered 102 SNPs that were significantly associated with end-of-induction MRD in both cohorts. Further functional analyses indicated that many of these host genetic variations were likely to influence treatment response via affecting host disposition of antineoplastic drugs.

Methods

Patients

Two cohorts of patients were included (Table 1S), with approval of the Institutional Review Board. From St. Jude Children's Research Hospital Total Therapy protocols XIIIB and XV, 371 children with newly diagnosed ALL had available germline DNA (i.e. collected at remission) and evaluable MRD status at the end of induction therapy. Of the ALL patients enrolled on the Children's Oncology Group (COG) P9906 study, 227 children had germline DNA and evaluable end-of-induction MRD status. The actual number of patients included in specific analyses is described below.

Treatment and MRD assessment

There were common and unique elements to the eligibility and treatment for the St. Jude and COG cohorts (Supplemental Fig. 1S), with details described elsewhere,[20,21] (http://www.acor.org/pedonc/diseases/ALLtrials/9906.html). Common elements included daily prednisone, weekly vincristine, weekly daunorubicin, thrice weekly asparaginase, and intrathecal therapy including methotrexate. After 28 days of therapy, St. Jude patients received additional therapy with cytarabine plus etoposide (Total XIIIB) or cytarabine plus cyclophosphamide and 6-mercaptopurine (Total XV). MRD was studied in bone marrow at days 19 and 46 by flow cytometry, with the latter time point corresponding to the end of induction treatment.[8] In contrast, COG patients finished the induction phase after 28 days of therapy, and MRD status was assessed using flow cytometry at day 8 (in blood) and at the end of the induction phase at day 28 (in bone marrow).[7] For St. Jude, MRD status was categorized as negative (<0.01%), positive (≥0.01%, but <1%), and high positive (≥ 1%). In COG, MRD classification was nearly identical: negative (≤0.01%), positive (>0.01%, but ≤1%), and high positive (>1%).

Diagnostic immunophenotype and molecular subtype analyses were performed as described.[7,8]

Genotyping, genotype imputation and quality control

DNA (500 ng) was digested with restriction enzymes (XbaI and Hind III for 100K SNP chip, and StyI and NspI for 500K SNP chip), amplified, labeled and hybridized to the Affymetrix GeneChip Human Mapping 100K and 500K Sets according to the manufacturer's instructions.

SNP genotypes were coded according to the number of B alleles in the genotype call as determined using BRLMM,[22] with the AA, AB, BB genotype calls coded as 0, 1, or 2, respectively. For genotypes that were not called by the BRLMM algorithm, we imputed the number of B alleles based on signal intensity and consistency with expected genotypes based on linkage disequilibrium,[23-25] whenever possible.

SNPs with minor allele frequency (MAF) < 1% or call rates < 95% (i.e. the number of samples with definitive genotype call at this SNP is <95% of the total number of samples typed for this SNP) were excluded (Fig. 1A); patient samples that failed to achieve 95% call rates (i.e. samples for which fewer than 95% of interrogated SNPs were successfully typed) were excluded (Fig. 1A and details in Supplemental Methods).

Genome-wide association analysis for MRD

MRD was treated as an ordinal variable, i.e. 1 for negative, 2 for positive, and 3 for high-positive, as defined above. In order to minimize confounding effects, patients with ALL subtypes (i.e. *E2A-PBX1*, *MLL* rearrangements, *BCR-ABL* ALL) that strongly predicted MRD and that differed in frequency between the two cohorts were excluded from the MRD analyses (Supplemental Table 1S). The final analysis included 476,796 SNPs, 318 St. Jude and 169 COG patients (Table 1 and Figure 1A).

SNPs associated with the end-of-induction MRD were identified by a three-step analysis (Figure 1B). Our goal was to find SNP genotypes that were associated with MRD in both cohorts—those that might be generalizable across treatment regimens for ALL. In step 1, we computed the statistical significance for each SNP genotype's association with MRD in each cohort separately. Rank (Spearman's) correlation was used for the test statistic, in order to account for both the ordinal nature of MRD and the gene dosage effect of genotypes. The P value was computed by a permutation-asymptotic hybrid method (see Supplemental Methods). An additive model was assumed, although the trend test is also reasonably robust to moderate

deviation from additivity.[26] In step 2, we determined the threshold for statistical significance by estimation of the false discovery rate (FDR) and an internal validation 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).[27] An internal validation (see Supplemental Methods) was then performed in each cohort. Based on the FDR estimates and the internal validation, a specific significance threshold ($P \leq 0.0125$) 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 ($P \leq 0.0125$) 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 (bidirectional validation) were prioritized for further bioinformatics and biological investigation, and analyses of association with additional relevant phenotypes.

Operating characteristics of the Spearman rank correlation test were determined via a simulation study (Supplemental Methods and Fig. 6S). The genotypes associated with MRD were also assessed by a pooled analysis that combined evidence across the two independent cohorts to provide a combined P-value for each SNP (Supplemental Methods). The FDR and the false positive report probability[28-32] (FPRP) for prioritized SNPs were estimated.

All statistical and computational analyses were performed using S-plus (Insightful Corp., Seattle, WA), R (www.r-project.org) and SAS (SAS Institute, Cary, NC).

Association of MRD SNPs with additional phenotypes

Antileukemic response: The relationship between the 102 overlapping MRD SNP genotypes and two additional leukemia response phenotypes was analyzed, in order to prioritize SNPs and to minimize the risk of false discoveries.

Patients were categorized into super responders, responders, and poor responders, based upon consideration of MRD status at two time points during the induction phase. MRD status was dichotomized as negative or positive, as defined above. Super responders were MRD-negative at the 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 Gray's test. Isolated central nervous system (CNS) relapse, isolated testicular relapse, combined CNS and testicular relapse, other relapse, lineage switch, second malignancy, and death in remission were incorporated in the analyses as competing events. Excluding individuals with *E2A-PBX1*, *MLL* rearrangements, or *BCR-ABL* 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 patients, 33 in St. Jude and 35 in COG experienced hematologic relapse (isolated and combined). St. Jude patients were divided into 4 strata according to their treatment protocol and risk classification, and COG patients formed the 5th stratum. Fine and Gray's cumulative incidence hazard regression model[33] was used to confirm the directional association with relapse for SNPs that achieved $P<0.1$ in the Gray test.

Pharmacokinetic studies: Three pharmacokinetic phenotypic data sets were available from a subset of St. Jude patients for antileukemic agents used during remission induction. Patients in these three 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 determined on day 29 of remission induction therapy, in 101 patients enrolled on St. Jude Total XIIIB.[34] 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 *CYP3A*[35] and P-glycoprotein,[36] a common mechanism of elimination that also affects prednisone,[34,37] vincristine,[38,39] and anthracyclines,[40,41] which were given to all patients in both cohorts.

The second data set included methotrexate plasma clearance in 319 patients treated on St. Jude Total XIIIB[20] and Total XV[21] protocols who received intravenous methotrexate as part of the early induction therapy. 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 and exert a systemic antileukemic effect.[42-44]

The third data set included intracellular methotrexate polyglutamate accumulation in ALL blasts at 44 hours after receiving up-front methotrexate in 330 patients treated on St. Jude trials.[45,46] Again, although intravenous methotrexate was given to and methotrexate polyglutamates were measured in only a subset of the St. Jude MRD cohort and none of the COG cohort, all patients in the MRD cohorts were exposed to methotrexate systemically via intrathecal injections.

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

Results

Identification and validation of genomic loci associated with end-of-induction MRD

A total of 588,920 SNPs were genotyped in germline DNA of 371 St. Jude and 227 COG patients. After quality control procedures were applied (Supplemental Methods and Table 1S), 476,796 SNPs were evaluated in 318 St. Jude and 169 COG patients (Fig. 1A and Table 1). We analyzed the association between germline SNP genotypes and MRD status independently in the St. Jude and COG cohorts (Fig. 1B). A P value threshold of 0.0125 was established based on false discovery rate (FDR) estimates and an internal validation inference (Supplemental Methods and Figure 2S). Through a rank-based bi-directional validation, a significant impact of germline variation on MRD identified in the St. Jude cohort was validated in the COG cohort ($P=2.2\times10^{-6}$), and that identified in the COG cohort was validated on the St. Jude cohort ($P<10^{-11}$) (Supplemental Methods).

In total, 102 SNPs exhibited significant concordant association with end-of-induction MRD ($P\leq0.0125$) in both the St. Jude and COG cohorts, with odds ratios ranging from 0.072 to 0.613 (median = 0.462) and from 1.63 to 7.42 (median = 2.18) (Supplemental Table 2S). Among these 102 SNPs, 50 were annotated to genes. Because 45 SNPs were clustered at 15 genomic loci by linkage disequilibrium (pair-wise $r^2>0.5$), these 102 SNPs represented 72 unique genomic loci (Supplemental Fig. 3S). A SNP in the *ST8SIA6* gene ($P=9.6\times10^{-8}$, combined cohort P value) had the strongest association with MRD but had no significant flanking SNPs and a relatively low MAF of 4% (Fig. 2, chromosome 10). The next highest ranked SNP (rs17007695) was in the *IL15* locus (Fig. 2, chromosome 4, Supplemental Table 2S) and was notable for strong ($P=8.8\times10^{-7}$, combined cohort P value) and comparable association with

MRD in both the St. Jude ($P=4.4\times10^{-4}$) and COG cohorts ($P=2.3\times10^{-4}$). Moreover, this SNP was flanked by four *IL15* SNPs (rs17015014, rs10519612, rs10519613, and rs35964658) that were also associated with MRD in both cohorts (Fig. 3A and Supplemental Table 2S), and these 5 SNPs were in linkage disequilibrium with each other (pair-wise r^2 from 0.48 to 0.97). Half of the St. Jude patients with the CC genotype, 35.6% of those with the CT genotype, and only 15.8% of patients with the TT genotype at the *IL15* SNP rs17007695 had detectable MRD at the end of induction therapy, with a similar trend observed in the COG cohort (Figure 3B). The CC genotype at *IL15* germline SNP rs17007695 was weakly associated (P=0.0701) with a higher *IL15* expression in leukemic blasts, and overexpression of *IL15* was associated with MRD in both cohorts (P=0.0342 in St. Jude and P=0.0035 in COG, Figure 4S).

All 102 SNPs remained significantly associated with MRD after adjustment for race, gender, leukocyte count at diagnosis, age and ALL subtypes (Supplemental Table 2S). To further explore possible confounding effects by race, we also examined the SNP vs. MRD associations in each major racial group. For instance, the GG genotype at rs13106616 was similarly associated with a lower risk of MRD across three race groups, although the allele frequency differed significantly by race (Supplemental Figure 5S). We also assessed the false positive report probability (FPRP) for these 102 SNPs and 82 (80.4%) exhibited FPRP<0.5 (Table 2S), a level associated with replicated associations in other contexts. [26,28-32]

Genome-wide association analysis for MRD using the 2-stage "discovery and validation" strategy—In addition to the bidirectional validation described above, 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 for each SNP genotype's association with MRD in the "discovery" cohort (St. Jude), estimating permutation-asymptotic hybrid P values for association with MRD as detailed in the Supplement. A P value threshold of 7×10^{-4} was arrived at by balancing the levels of false negative and false positive errors using the profile information criterion (Supplemental Figure 7S);[27] 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 $P \leq 0.05$, significantly more than what would be expected by chance (P=0.021, Fisher's exact test), and these are highlighted in Supplemental Table 2S. When the P value threshold was set at 0.0125 for the discovery cohort (St. Jude), 8635 SNPs met this cutoff, 330 of which were validated in the COG cohort with $P \leq 0.05$, exceeding what would be expected by chance ($P=1.8\times10^{-9}$, Fisher's exact test).

Relation of MRD-associated SNPs to other antileukemic response phenotypes

Although end-of-induction therapy MRD is highly predictive of long-term treatment outcome, the early reduction of leukemic burden during therapy is also informative.[47] 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 (super responders, 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<0.05) with early response in both cohorts (Supplemental Table 3S).

Of the 102 SNPs, 21 were significantly associated with hematologic relapse (P<0.05) by stratified Gray's test and in a cumulative incidence hazard regression model (P<0.05). For instance, there was a monotonic relationship between the number of copies of the C allele at

rs1486649 (an intergenic SNP) and the risk of hematologic relapse (Figure 4A and B, Supplemental Table 3S).

Relation of MRD-associated SNPs with 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 (Supplemental Table 3S). In total, 21 of the 102 MRD-related SNPs exhibited significant association with antileukemic agent pharmacokinetics, with 3 SNPs predicting more than one pharmacokinetic phenotype. Eight of 102 SNPs were associated with clearance of methotrexate (at $P<0.05$); all 8 genotypes associated with positive MRD and greater drug clearance. Ten of the 102 SNPs were associated with the pharmacokinetics of etoposide, with 7 of 10 associating 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 associating with positive MRD and lower methotrexate polyglutamates. Thus, of 24 significant associations, 20 were directly consistent with a pharmacokinetically intuitive association with MRD, i.e. lower drug exposure translated into a higher level of MRD. Specific genotypes linked higher methotrexate clearance (decreased drug exposure) (Fig. 5A), lower accumulation of methotrexate polyglutamates in the leukemic blasts (Fig. 5B), and greater clearance of etoposide (Fig. 5C) with a higher frequency of MRD.

Comment

Eradication of malignant cells by chemotherapy is a composite phenotype which depends not only on the somatically acquired characteristics of the malignant cells but also upon inherent patient characteristics. Childhood ALL has long served as a prototype for a malignancy that is curable with drugs. Early assessments of MRD strongly predict cure rates, and are used to modify therapy.[3-9,48] Eradication of MRD is affected by genetic characteristics of the blasts (e.g. the presence of the Philadelphia chromosome or the *TEL/AML1* translocation) and by host characteristics such as age.[7,8] Using a candidate gene approach, a few germline genetic variations have been shown to affect the level of MRD,[16,49] but this has not been previously assessed on a genome-wide level. Herein, we used an agnostic genome-wide interrogation to identify 102 germline genetic variations that affected the level of residual leukemia in two independent cohorts of patients, and found that a high proportion (63 of 102 SNPs or 61.7%) also affected early response, relapse risk, or antileukemic drug disposition.

One of the strongest signals from the genome-wide scan came from 5 SNPs located in and around the *IL15* gene, a proliferation-stimulatory cytokine.[50,51] IL15 can protect lymphoid tumors from glucocorticoid-induced apoptosis *in vitro*,[52] and *IL15* expression in ALL blasts has been linked to both CNS involvement at diagnosis and an increased risk of CNS relapse.[53] Both higher *IL15* gene expression ($P=0.0342$ in St. Jude and $P=0.003$ in COG) and germline SNP genotypes were associated with an increased risk of positive MRD at the end of induction therapy (Supplemental Figure 4S), and we found a trend ($P=0.0701$) towards a significant relationship between *IL15* germline SNP genotypes and *IL15* gene expression in ALL leukemic blasts. Several of the *IL15* SNPs that predicted MRD have been associated with enhanced *IL15* transcription/translation efficiency *in vitro*.[54] 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 two independently-treated cohorts points to its importance in determining ALL response, either as a prognostic marker or as a therapeutic target.

As 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

three relevant host pharmacokinetic phenotypes: systemic clearance of two antileukemic agents (etoposide and methotrexate) and intracellular disposition of the latter. Although 4-8 different antileukemic agents were used in these two cohorts, remarkably, 21 of the 102 MRD-predicting SNPs we identified were also significantly associated with disposition of these two 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-predicting 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 cytochrome *P4503A* 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.[34,37-41] The majority (83.3%) of the associations between SNP genotypes and drug disposition were pharmacologically intuitive, with genotypes that predicted increased drug exposure linked to lower levels of MRD. Together, these results suggest that more attention be given to details of drug administration and risk factors for rapid drug clearance, in addition to the considerable attention already placed upon better risk classification of ALL to tailor therapy intensity.

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 post-remission 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 upon therapy. For this reason, we purposefully chose two 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. The advantage of our bi-directional statistical approach is that the SNPs we identified may be more likely to have external validity for other patient groups; the disadvantage is that we might have missed SNPs more specific to the few elements of therapy that differed between the cohorts.

It is important to consider race, both from the standpoint of its possible effects on antileukemic drug efficacy[55-57] and from its influence on germline SNP allele frequency.[58] The influence of race on ALL cure rates may be due to differences among races in the delivery of care, patient compliance, frequencies of poor-prognosis ALL subtypes, or to differences in allele frequencies for germline polymorphisms.[16] 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 (Supplemental Table 2S). Thus, population stratification was unlikely to have affected the SNP genotype/phenotype associations we discovered, consistent with other recent studies.[59,60] The fact that SNP genotypes maintained their significance after adjusting for race, despite some cases of substantial differences in allele frequency by race (Supplemental Figure 5S), 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 two 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 (e.g. *CYP2C9/VKORC* for warfarin[61,62] and *TPMT* for

thiopurine[49,63]) may have effect sizes that exceed those likely to be observed for multigenic common diseases (e.g. diabetes and arthritis),[24] and therefore smaller sample size may suffice in the former. By identifying 102 SNPs based on evidence of association in two independent cohorts, and also by further validation of 62% of these SNPs (Supplemental Table 3S) as associating with the related phenotypes of relapse, “super response” at days 8 or 19, and antileukemic drug pharmacokinetics, we have further decreased the chance for false discoveries. The SNPs we identified may be in linkage disequilibrium[64] with the truly causative genetic variants that have not yet been interrogated directly by our genotyping platform (Supplemental Table 4S). Importantly, few of the 102 polymorphisms we identified have previously been suggested as candidates for affecting anticancer 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We are indebted to all patients and their parents who participated St. Jude Total Therapy protocols XIIIB and XV and COG P9906 study, clinicians and research staff at St. Jude and COG institutions, Drs. Jeannette Pullen and Andrew Carroll for assistance in classification of patients with ALL, Tianhe Zhang for his help in data analysis.

Funding/Support: This work was supported by CA093552-02, NCI CA 51001, CA 78224, CA21765, CA R37 36401, and the NIH/NIGMS Pharmacogenetics Research Network and Database (U01 GM61393, U01GM61374 http://www.pharmgkb.org) from the National Institutes of Health; American Lebanese Syrian Associated Charities (ALSAC); and by CureSearch.

Reference List

(1). Pui CH, Evans WE. Treatment of acute lymphoblastic leukemia. N Engl J Med 2006;354(2):166–178. [PubMed: 16407512]

(2). Pui CH, Jeha S. New therapeutic strategies for the treatment of acute lymphoblastic leukaemia. Nat Rev Drug Discov 2007;6(2):149–165. [PubMed: 17268486]

(3). van Dongen JJ, Seriu T, Panzer-Grumayer ER, et al. Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. Lancet 1998;352(9142):1731–1738. [PubMed: 9848348]

(4). Flohr T, Schrauder A, Cazzaniga G, et al. Minimal residual disease-directed risk stratification using real-time quantitative PCR analysis of immunoglobulin and T-cell receptor gene rearrangements in the international multicenter trial AIEOP-BFM ALL 2000 for childhood acute lymphoblastic leukemia. Leukemia. 2008

(5). Szczepanski T, Orfao A, van dV V, Miguel JF San, van Dongen JJ. Minimal residual disease in leukaemia patients. Lancet Oncol 2001;2(7):409–417. [PubMed: 11905735]

(6). Cave H, Bosch van der Werff ten, Suciu S, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia. European Organization for Research and Treatment of Cancer--Childhood Leukemia Cooperative Group. N Engl J Med 1998;339(9):591–598. [PubMed: 9718378]

(7). Borowitz MJ, Pullen DJ, Shuster JJ, et al. Minimal residual disease detection in childhood precursor-B-cell acute lymphoblastic leukemia: relation to other risk factors. A Children's Oncology Group study. Leukemia 2003;17(8):1566–1572. [PubMed: 12886244]

(8). Coustan-Smith E, Sancho J, Hancock ML, et al. Clinical importance of minimal residual disease in childhood acute lymphoblastic leukemia. Blood 2000;96(8):2691–2696. [PubMed: 11023499]

(9). Zhou J, Goldwasser MA, Li A, et al. Quantitative analysis of minimal residual disease predicts relapse in children with B-lineage acute lymphoblastic leukemia in DFCI ALL Consortium Protocol 95-01. Blood 2007;110(5):1607–1611. [PubMed: 17485550]

(10). Carroll WL, Bhojwani D, Min DJ, Moskowitz N, Raetz EA. Childhood acute lymphoblastic leukemia in the age of genomics. Pediatr Blood Cancer 2006;46(5):570–578. [PubMed: 16365862]

(11). Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell 2002;1(2):133–143. [PubMed: 12086872]

(12). Holleman A, den Boer ML, de Menezes RX, et al. The expression of 70 apoptosis genes in relation to lineage, genetic subtype, cellular drug resistance, and outcome in childhood acute lymphoblastic leukemia. Blood 2006;107(2):769–776. [PubMed: 16189266]

(13). Lugthart S, Cheok MH, den Boer ML, et al. Identification of genes associated with chemotherapy crossresistance and treatment response in childhood acute lymphoblastic leukemia. Cancer Cell 2005;7(4):375–386. [PubMed: 15837626]

(14). Schmidt S, Rainer J, Riml S, et al. Identification of glucocorticoid-response genes in children with acute lymphoblastic leukemia. Blood 2006;107(5):2061–2069. [PubMed: 16293608]

(15). Tissing WJ, den Boer ML, Meijerink JP, et al. Genomewide identification of prednisolone-responsive genes in acute lymphoblastic leukemia cells. Blood 2007;109(9):3929–3935. [PubMed: 17218380]

(16). Davies SM, Borowitz MJ, Rosner GL, et al. Pharmacogenetics of minimal residual disease response in children with Acute Lymphoblastic Leukemia (ALL): a report from the Children's Oncology Group. Blood. 2008

(17). Rocha JC, Cheng C, Liu W, et al. Pharmacogenetics of outcome in children with acute lymphoblastic leukemia. Blood 2005;105(12):4752–4758. [PubMed: 15713801]

(18). Relling MV, Hancock ML, Boyett JM, Pui CH, Evans WE. Prognostic importance of 6-mercaptopurine dose intensity in acute lymphoblastic leukemia. Blood 1999;93(9):2817–2823. [PubMed: 10216075]

(19). Fleury I, Primeau M, Doreau A, et al. Polymorphisms in genes involved in the corticosteroid response and the outcome of childhood acute lymphoblastic leukemia. Am J Pharmacogenomics 2004;4(5):331–341. [PubMed: 15462611]

(20). Pui CH, Sandlund JT, Pei D, et al. Improved outcome for children with acute lymphoblastic leukemia: results of Total Therapy Study XIIIB at St Jude Children's Research Hospital. Blood 2004;104(9):2690–2696. [PubMed: 15251979]

(21). Pui CH, Relling MV, Sandlund JT, Downing JR, Campana D, Evans WE. Rationale and design of Total Therapy Study XV for newly diagnosed childhood acute lymphoblastic leukemia. Ann Hematol 2004;83(Suppl 1):S124–S126. [PubMed: 15124703]

(22). Rabbee N, Speed TP. A genotype calling algorithm for affymetrix SNP arrays. Bioinformatics 2006;22(1):7–12. [PubMed: 16267090]

(23). Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 2006;78(4):629–644. [PubMed: 16532393]

(24). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447(7145):661–678. [PubMed: 17554300]

(25). Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet 2008;40(2):161–169. [PubMed: 18193043]

(26). McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 2008;9(5):356–369. [PubMed: 18398418]

(27). Cheng C, Pounds SB, Boyett JM, Pei D, Kuo ML, Roussel MF. Statistical significance threshold criteria for analysis of microarray gene expression data. Stat Appl Genet Mol Biol 2004;3Article36

(28). Wacholder S, Chanock S, Garcia-Closas M, El GL, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst 2004;96 (6):434–442. [PubMed: 15026468]

(29). Hunter DJ, Kraft P. Drinking from the fire hose--statistical issues in genomewide association studies. N Engl J Med 2007;357(5):436–439. [PubMed: 17634446]

(30). Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. N Engl J Med 2007;357(5):443–453. [PubMed: 17634449]

(31). Richards JB, Rivadeneira F, Inouye M, et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. Lancet 2008;371(9623):1505–1512. [PubMed: 18455228]

(32). Valdes AM, Loughlin J, Timms KM, et al. Genome-wide Association Scan Identifies a Prostaglandin-Endoperoxide Synthase 2 Variant Involved in Risk of Knee Osteoarthritis. Am J Hum Genet 2008;82(6):1231–1240. [PubMed: 18471798]

(33). Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association 1999;94(446):496–509.

(34). Kishi S, Yang W, Boureau B, et al. Effects of prednisone and genetic polymorphisms on etoposide disposition in children with acute lymphoblastic leukemia. Blood 2004;103(1):67–72. [PubMed: 12969965]

(35). Relling MV, Nemec J, Schuetz EG, Schuetz JD, Gonzalez FJ, Korzekwa KR. O-demethylation of epipodophyllotoxins is catalyzed by human cytochrome P450 3A4. Mol Pharmacol 1994;45(2): 352–358. [PubMed: 8114683]

(36). Leu BL, Huang JD. Inhibition of intestinal P-glycoprotein and effects on etoposide absorption. Cancer Chemother Pharmacol 1995;35(5):432–436. [PubMed: 7850926]

(37). Yates CR, Chang C, Kearbey JD, et al. Structural determinants of P-glycoprotein-mediated transport of glucocorticoids. Pharm Res 2003;20(11):1794–1803. [PubMed: 14661924]

(38). Villikka K, Kivisto KT, Maenpaa H, Joensuu H, Neuvonen PJ. Cytochrome P450-inducing antiepileptics increase the clearance of vincristine in patients with brain tumors. Clin Pharmacol Ther 1999;66(6):589–593. [PubMed: 10613614]

(39). Plasschaert SL, Groninger E, Boezen M, et al. Influence of functional polymorphisms of the MDR1 gene on vincristine pharmacokinetics in childhood acute lymphoblastic leukemia. Clin Pharmacol Ther 2004;76(3):220–229. [PubMed: 15371983]

(40). Lewis AD, Lau DH, Duran GE, Wolf CR, Sikic BI. Role of cytochrome P-450 from the human CYP3A gene family in the potentiation of morpholino doxorubicin by human liver microsomes. Cancer Res 1992;52(16):4379–4384. [PubMed: 1643634]

(41). van AJ, van TO, Beijnen JH. The role of mdr1a P-glycoprotein in the biliary and intestinal secretion of doxorubicin and vinblastine in mice. Drug Metab Dispos 2000;28(3):264–267. [PubMed: 10681369]

(42). Bleyer WA, Nelson JA, Kamen BA. Accumulation of methotrexate in systemic tissues after intrathecal administration. J Pediatr Hematol Oncol 1997;19(6):530–532. [PubMed: 9407940]

(43). Bostrom BC, Erdmann GR, Kamen BA. Systemic methotrexate exposure is greater after intrathecal than after oral administration. J Pediatr Hematol Oncol 2003;25(2):114–117. [PubMed: 12571461]

(44). Conter V, Arico M, Valsecchi MG, et al. Extended intrathecal methotrexate may replace cranial irradiation for prevention of CNS relapse in children with intermediate-risk acute lymphoblastic leukemia treated with Berlin-Frankfurt-Munster-based intensive chemotherapy. The Associazione Italiana di Ematologia ed Oncologia Pediatrica. J Clin Oncol 1995;13(10):2497–2502. [PubMed: 7595699]

(45). Kager L, Cheok M, Yang W, et al. Folate pathway gene expression differs in subtypes of acute lymphoblastic leukemia and influences methotrexate pharmacodynamics. J Clin Invest 2005;115 (1):110–117. [PubMed: 15630450]

(46). Masson E, Relling MV, Synold TW, et al. Accumulation of methotrexate polyglutamates in lymphoblasts is a determinant of antileukemic effects in vivo. A rationale for high-dose methotrexate. J Clin Invest 1996;97(1):73–80. [PubMed: 8550853]

(47). Schultz KR, Pullen DJ, Sather HN, et al. Risk- and response-based classification of childhood B-precursor acute lymphoblastic leukemia: a combined analysis of prognostic markers from the

Pediatric Oncology Group (POG) and Children's Cancer Group (CCG). Blood 2007;109(3):926–935. [PubMed: 17003380]

(48). Coustan-Smith E, Behm FG, Sanchez J, et al. Immunological detection of minimal residual disease in children with acute lymphoblastic leukaemia. Lancet 1998;351(9102):550–554. [PubMed: 9492773]

(49). Stanulla M, Schaeffeler E, Flohr T, et al. Thiopurine methyltransferase (TPMT) genotype and early treatment response to mercaptopurine in childhood acute lymphoblastic leukemia. JAMA 2005;293(12):1485–1489. [PubMed: 15784872]

(50). Budagian V, Bulanova E, Paus R, Bulfone-Paus S. IL-15/IL-15 receptor biology: a guided tour through an expanding universe. Cytokine Growth Factor Rev 2006;17(4):259–280. [PubMed: 16815076]

(51). Fehniger TA, Caligiuri MA. Interleukin 15: biology and relevance to human disease. Blood 2001;97(1):14–32. [PubMed: 11133738]

(52). Tinhofer I, Marschitz I, Henn T, Egle A, Greil R. Expression of functional interleukin-15 receptor and autocrine production of interleukin-15 as mechanisms of tumor propagation in multiple myeloma. Blood 2000;95(2):610–618. [PubMed: 10627470]

(53). Cario G, Izraeli S, Teichert A, et al. High interleukin-15 expression characterizes childhood acute lymphoblastic leukemia with involvement of the CNS. J Clin Oncol 2007;25(30):4813–4820. [PubMed: 17947730]

(54). Zhang XJ, Yan KL, Wang ZM, et al. Polymorphisms in interleukin-15 gene on chromosome 4q31.2 are associated with psoriasis vulgaris in Chinese population. J Invest Dermatol 2007;127(11):2544–2551. [PubMed: 17554368]

(55). Carroll WL. Race and outcome in childhood acute lymphoblastic leukemia. JAMA 2003;290(15):2061–2063. [PubMed: 14559962]

(56). Kadan-Lottick NS, Ness KK, Bhatia S, Gurney JG. Survival variability by race and ethnicity in childhood acute lymphoblastic leukemia. JAMA 2003;290(15):2008–2014. [PubMed: 14559954]

(57). Pui CH, Sandlund JT, Pei D, et al. Results of therapy for acute lymphoblastic leukemia in black and white children. JAMA 2003;290(15):2001–2007. [PubMed: 14559953]

(58). Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38(8):904–909. [PubMed: 16862161]

(59). Millikan RC. Re: Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst 2001;93(2):156–158. [PubMed: 11208892]

(60). Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst 2000;92(14):1151–1158. [PubMed: 10904088]

(61). Schwarz UI, Ritchie MD, Bradford Y, et al. Genetic determinants of response to warfarin during initial anticoagulation. N Engl J Med 2008;358(10):999–1008. [PubMed: 18322281]

(62). Aithal GP, Day CP, Kesteven PJ, Daly AK. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. Lancet 1999;353(9154):717–719. [PubMed: 10073515]

(63). Evans WE, Hon YY, Bomgaars L, et al. Preponderance of thiopurine S-methyltransferase deficiency and heterozygosity among patients intolerant to mercaptopurine or azathioprine. J Clin Oncol 2001;19(8):2293–2301. [PubMed: 11304783]

(64). Malo N, Libiger O, Schork NJ. Accommodating linkage disequilibrium in genetic-association analyses via ridge regression. Am J Hum Genet 2008;82(2):375–385. [PubMed: 18252218]

Table 1

Patient Characteristics and Relation to Minimal Residual Disease.

		St.Jude		COG	
		Patients (%) n=318	**Relation to MRD (P value) ****	**Patients (%) n=169**	**Relation to MRD (P value) ****
Race *	Caucasian	246 (77%)		110 (65%)	
	African	45 (14%)		8 (5%)	
	Other	27 (9%)	0.63	51 (30%)	0.179
Gender	Male	181 (57%)		118 (70%)	
	Female	137 (43%)	0.33	51 (30%)	0.132
Age at Diagnosis	<1 year	2 (1%)		0	
	1-10 years	236 (74%)		51 (30%)	
	>10 years	80 (25%)	0.28	118 (70%)	0.057
WBC at Diagnosis	<50k cells/ul	237 (75%)		94 (56%)	
	>50k cells/ul	81 (25%)	0.067	75 (44%)	0.155
Lineage	B-lineage	248 (78%)		169 (100%)	
	T-cell	70(22%)	0.062	0	NA
Molecular Subtype	*TEL-AML1*	61 (19%)	0.33	3 (2%)	0.45
	BCR-ABL	0		0	
	E2A-PBX1	0		0	
	MLL	0		0	
	rearrangements No Common Translocation	257 (81%)	NA	166 (98%)	NA
End-of-induction MRD	<0.01%	257 (81%)		111 (66%)	
	0.01-1%	53 (17%)		36 (21%)	
	>1%	8 (2%)		22 (13%)	

WBC: white blood cell count; MRD: minimal residual disease, NA: non-applicable

JAMA. Author manuscript; available in PMC 2009 July 28.

[*] Race was assigned based on germline genotype of ~600,000 SNPs, as described in the Supplemental Information.

[**] Association between MRD (treated as a categorical variable, as defined in the Methods) and patient characteristics was assessed using a χ^2 test.