Comprehensive genetic analysis of cytarabine sensitivity in a cell-based model identifies polymorphisms associated with outcome in AML patients

Eric R. Gamazon¹, Jatinder Lamba², Stanley Pounds³, Amy L. Stark¹, Heather E. Wheeler¹, Xueyuan Cao³, Hae K. Im⁴, Amit K. Mitra², Jeffrey E. Rubnitz⁵, Raul C. Ribeiro⁵, Susana Raimondi⁶, Dario Campana⁵,⁶, Kristine R. Crews⁷, Shan S. Wong¹, Marleen Welsh¹, Imge Hulur¹, Lidija Gorsic¹, Christine M. Hartford⁸, Wei Zhang⁹, Nancy J. Cox¹, M. Eileen Dolan¹∗

¹Department of Medicine and ⁴Department of Health Studies, The University of Chicago, Chicago, IL 60637; ²Department of Experimental & Clinical Pharmacology, College of Pharmacy, PUMA-Institute of Personalized Medicine, University of Minnesota, MN; ³Department of Biostatistics, ⁵Department of Oncology, ⁶Department of Pathology, ⁷Department of Pharmaceutical Sciences, ⁸Department of Bone Marrow Transplantation and Cellular Therapy, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, ⁹Department of Pediatrics and Institute of Human Genetics, University of Illinois, Chicago, IL 60612, USA
*Address for correspondence and reprints: M. Eileen Dolan, 900 E. 57th St, KCBD Rm 7100, The University of Chicago, Chicago, IL 60637. Phone: (773) 702-4441; Fax: (773) 702-9268; E-mail: edolan@medicine.bsd.uchicago.edu

Key words: Pharmacogenomics, GWAS, meta-analysis, susceptibility to chemotherapy

Abbreviations: AML, acute myeloid leukemia; cytarabine, cytarabine arabinoside; DCK, deoxycytidine kinase; NT5C2, 5’- nucleotidase cytosolic II; ara-CTP, araC-triphosphate; CDA, cytidine deaminase; hENT1, human equilibrative nucleoside transporter 1; dCTP, deoxycytidine triphosphate; LCLs, lymphoblastoid cell lines; SNPs, single nucleotide polymorphisms; GWAS, genome-wide association studies; MCC, mutated in colorectal cancer; QTDT, quantitative transmission disequilibrium test; eQTL, expression quantitative trait loci; LD, linkage disequilibrium; MAF, minor allele frequency; CEU refers to the HapMap samples of Northern and Western European descents; YRI refers to the Yoruba HapMap samples from Ibadan, Nigeria; ASW refers to African ancestry in Southwest USA; ASN refers to Han Chinese in Beijing, China and Japanese in Tokyo, Japan; TRM, treatment-related mortality; RFS, relapse-free survival, EFS, event-free survival; MRD, minimal residual disease; OS, overall survival
Key Points

- *Preclinical cell-based model identifies SNPs associated with cytarabine sensitivity that also associate with outcome in leukemia patients.*

- *SNPs within MCC gene were associated with cytarabine sensitivity in lymphoblastoid cell lines and leukemic blasts from patients.*

Abstract

Cytarabine is used to treat acute myeloid leukemia (AML) and resistance to therapy is a common reason for treatment failure. We used a whole-genome approach to investigate the genetic determinants of cytarabine-induced cytotoxicity. We performed a meta-analysis of genome-wide association studies involving 523 lymphoblastoid cell lines (LCLs) from individuals of European, African, Asian and African American ancestry. Several of the highest-ranked SNPs were within the “mutated in colorectal cancers” (MCC) gene. MCC expression was induced by cytarabine treatment from 1.7- to 26.6-fold in LCLs. Thirty three SNPs ranked at the top of the meta-analysis (p < 10^-5) were successfully tested in a clinical trial of patients randomized to receive low-dose or high-dose cytarabine plus daunorubicin and etoposide; of these, 18 showed association (p < 0.05) with either cytarabine IC50 in leukemia cells or clinical response parameters (minimal residual disease, overall survival, treatment-related mortality). This count (n=18) was significantly greater than expected by chance (p=0.016). For rs1203633, LCLs with AA genotype were more sensitive to cytarabine-induced cytotoxicity (p = 1.31 x 10^-6), and AA (vs GA or GG) genotype was associated with poorer overall survival (p = 0.015) likely a result of greater treatment-related mortality (p = 0.0037) in AML patients.

(Multicenter AML02 study trial: NCT00136084)
Introduction

Acute myeloid leukemia (AML) is the most common form of acute leukemia in adults and also occurs in children. Despite the genetic heterogeneity of the disease, patients have been treated for decades with similar combinations of cytarabine and anthracyclines with little improvement in overall survival. Although the majority of patients (50–60%) under 60 years achieve complete remission with traditional anthracycline- and cytarabine-based induction regimens, the long-term survival rates continue to be around 30% to 40% for adults and 60% for children. Outcomes are worse for patients ≥ 60 years, with complete response rates in the range of 40 to 55% and poor long-term survival rates. The main reason for treatment failure among patients with AML is resistance to therapy. In addition, treatment with cytarabine is associated with a number of adverse side effects including myelosuppression, infections, mucositis, neurotoxicity and acute pulmonary syndrome.

Cytarabine requires activation through intracellular phosphorylation to araC-triphosphate (ara-CTP). The mechanism of action of cytarabine involves incorporation of ara-CTP in place of deoxycytidine triphosphate (dCTP) resulting in chain termination, blocking DNA and RNA synthesis and causing leukemic cell death. One of the greatest predictors of response to cytarabine is the intracellular concentration of ara-CTP ex vivo and in circulating blasts of patients. Resistance is likely due to inefficient uptake of araCTP, reduced levels of deoxycytidine kinase (DCK), increased levels of deactivating enzymes 5’- nucleotidase (NT5C2) or cytidine deaminase (CDA), or increased cellular dCTP pools that compete with ara-CTP for incorporation in DNA.
Candidate approaches and genome-wide association studies (GWAS) have been used to identify genetic variables that are important in inter-individual variability in sensitivity to cytarabine. Candidate gene studies revealed that genes within the cytarabine pharmacokinetic pathway including DCK, CDA, NT5C2, NT5C3 and human equilibrative nucleoside transporter 1 (hENT1) contribute to sensitivity to cytarabine.

In previous work, we found genetic variants associated with cytarabine sensitivity in 85 European (CEU) and 89 African (YRI) LCLs that were specific to each population (505 SNPs for CEU and 397 SNPs for YRI at p < 1 x 10⁻⁴, no overlap) as well as associated variants in the “African American from the Southwestern United States” (ASW) population. The results of these cell-based models can be used in conjunction with clinical trials for discovery of SNPs associated with chemotherapeutic sensitivity, given the challenge of accruing large patient cohorts receiving the same drug regimen for discovery and replication GWAS in oncology. In this study, our goal was to identify variants that associate with cytarabine-induced cytotoxicity in 523 LCLs from different world populations representing European, African, Asian, and African American ancestries providing a robust set of SNPs for studies in clinical trials. We evaluated the significance of the most highly ranked SNPs (p < 10⁻⁵) with cytarabine-induced apoptosis and with treatment outcome in AML patients who received cytarabine-containing therapy.

Material and Methods

Meta-analysis of Six GWA Studies

Details on the cytotoxicity assays in LCLs and the GWAS in individual panels are found in the Supplementary Materials. To determine SNP associations with cytarabine-
induced cytotoxicity across populations, we conducted a meta-analysis on the results of
the individual GWAS from the 6 panels using METAL, which combines SNP p-values
across studies taking into account a study-specific weight (sample size) and direction of
effect (positive or negative beta) 19. Z-scores, derived from the p-values for each SNP,
were combined across studies in a weighted sum, wherein the weights were defined to be
proportional to the square-root of the sample size for each study 19. The Q-Q plot of the
corresponding p-values was generated using R. Local region plots of top associated SNPs
were generated by LocusZoom 20.

Association of Top Meta-analysis SNPs with Apoptosis

Apoptosis may underlie cytarabine-induced cytotoxicity; thus, the top thirty seven
SNPs (p < 1 x 10^-5) selected from the meta-analysis of cytotoxicity in LCLs were tested
for association with cytarabine-induced apoptosis. The apoptotic effects of cytarabine
were determined in CEU1/2, YRI1/2, ASW and Han Chinese in Beijing, China (CHB)
samples. Apoptosis, as measured by caspase 3/7 activation induced by 40 µM cytarabine
24 hours post drug treatment, was measured as previously described 21. Association
analyses with apoptosis were done within each panel; results were then combined into a
single p-value for each SNP using the meta-analysis method.

Clinical Samples

Details of the multicenter AML02 study trial (NCT00136084) protocol and
outcome are described elsewhere 2. Briefly, from October 13, 2002, to June 19, 2008, 232
children with de-novo AML (n=206), therapy-related or myelodysplastic syndromes
(MDS)-related AML (n=12), or mixed-lineage leukemia (n=14) were randomized to receive high-dose cytarabine (3 g/m² intravenously over 3 h, given every 12 h on days 1, 3 and 5, n=113) or low-dose cytarabine (100 mg/m² intravenously over 30 min, given every 12 h on days 1–10, n=11) plus daunorubicin (50 mg/m² intravenously over 6 h on days 2, 4 and 6) and etoposide (100 mg/m² intravenously over 4 h on days 2–6). All patients were chemo-naïve at the time of enrollment except for four subjects who had AML as a second malignancy. Patients were randomly assigned to receive high-dose or low-dose cytarabine. The patient population included 69.6% Caucasian, 18.7% African American, and 11.7% with other ethnic backgrounds. Primary bone marrow samples were obtained after informed consent was obtained from them or from their parents/guardians, with assent from the patients, as appropriate, in accordance with the Declaration of Helsinki. This study and use of these samples were approved by the institutional review board at St. Jude Children's Research Hospital.

The top thirty seven SNPs selected from the meta-analysis in LCLs were genotyped in genomic DNA from these AML patients using a sequenom (iPLEX) mass spectrometry-based multiplex genotyping assay (Sequenom, Inc San Diego, CA) at BioMedical Genomics Center, University of Minnesota. Of the 37 SNPs, 34 SNPs were successfully genotyped (3 failed genotyping including one within MCC) and one was not polymorphic in AML patients.

Minimal residual disease (MRD), relapse-free survival (RFS), event free survival (EFS), overall survival (OS) and treatment-related mortality (TRM) were determined as previously described ². Ex vivo sensitivity to cytarabine (IC₅₀) of leukemic cells obtained at diagnosis was determined in patients enrolled on the AML02 protocol using the 4-day
MTT cytotoxicity assay as described previously\textsuperscript{22,23}. The statistical analysis of clinical samples is described in Supplemental Materials.

\textit{Evaluation of Meta-analysis LCL SNPs in Patients}

We tested the 33 SNPs that had an association with cytarabine-induced cytotoxicity in LCLs (p<1 x 10\textsuperscript{-5}) for association (p<0.05 and concordant direction) with either cytarabine IC\textsubscript{50} in leukemia cells or with clinical response parameters (MRD on day 22, OS, or TRM). We refer to these SNPs as “clinically validated SNPs”.

For assessment of the statistical significance of the number of clinically validated SNPs, patient label was permuted (n=5000) in the Caucasian patients while preserving the genotype correlation structure among the 33 SNPs and the correlation structure among the clinical traits tested. In each permuted dataset, the association tests with the clinical phenotypes were performed; SNPs with p<0.05 were selected. A permuted dataset was considered a “success” if the number of (unique) selected SNPs from the association tests was equal to or greater than the number of clinically validated SNPs. Empirical significance of the excess of clinically validated SNPs was estimated as the proportion of such “successes”.

\textit{MCC Gene Expression Post-Cytarabine Treatment}

Quantitative real-time polymerase chain reaction (qRT-PCR) was performed to measure the level of expression of MCC over time following cytarabine treatment. Details are found in Supplementary Materials.
RESULTS

Cellular Sensitivity Phenotypes

Previously, genetic variants associated with cytarabine-induced cytotoxicity were identified in 85 CEU, 89 YRI, 83 ASW from the International HapMap lymphoblastoid cell line collection. In the present study, we extended these studies to increase power to detect associations and to identify cross-population SNPs. Prior to performing a meta-analysis, we evaluated an additional 89 CEU (174 total), 87 YRI (176 total), and 90 ASN (consisting of 45 JPT and 45 CHB) for cellular sensitivity to cytarabine using a short-term growth inhibition assay. We determined the percent survival at five different concentrations for each cell line, from which AUC was calculated. Figure 1 illustrates the box plot of the distribution of log₂ AUC values in the 4 populations of cell lines. The log₂ AUC values in each cell line panel showed no departure from normality, with the exception of the ASW panel, which was thus rank-transformed to normality (Supplementary Figure 1).

Meta-analysis across population panels identifies SNPs associated with cytarabine-induced cytotoxicity

Using quantitative trait disequilibrium total association (QTDT), genome-wide association studies were performed in each panel of LCLs to identify SNP associations with cellular sensitivity to cytarabine. To increase the density of interrogated SNPs, ungenotyped markers in Phase III of the CEU, YRI, and ASW samples were imputed using BEAGLE. To increase statistical power, we conducted a meta-analysis on the results of the individual GWAS of each of the six panels. The meta-analysis combined
the results of individual GWA studies from each panel to identify SNPs associated with
cytarabine AUC across populations, accounting for sample size and direction of effect.
Approximately 60% of the most highly ranked SNPs (p < 10^{-4}, n=370) had consistent
allelic direction of effect in all 6 panels, as perhaps expected, since such SNPs would rise
to the top of the meta-analysis results. Figure 2A illustrates a Manhattan plot
summarizing the results of the meta-analysis. Table 1 lists the top 37 SNPs (p < 10^{-5}),
their location, p-value and the directionality in each panel evaluated (ASN, CEU1/2,
CEU3, ASW, YRI1/2, YRI3). While most of these SNPs were common (MAF > 5%) in
all the panels, several SNPs were specific to a population class (e.g., monomorphic in the
other populations) and therefore were not interrogated in as many individuals (Table 1,
column heading: weight).

To ascertain the significance of SNPs within known candidate genes in the
cytarabine pathway (CDA, NT5C2, NT5C3, SLC29A1, and DCK)\textsuperscript{24}, we retrieved the
meta-analysis p-value for each of the SNPs located within the pathway and polymorphic
at MAF>5% in any of the populations examined here (Supplementary Table 1). Several
of the SNPs in 3 genes (CDA, NT5C2, and SLC29A1) showed nominal significance in our
LCL model (p<0.05).

Within the context of an unbiased genome-wide approach, we were also interested
in identifying novel susceptibility loci outside the known cytarabine pathway. While no
SNP reached genome-wide significance (Bonferroni) from the meta-analysis, we found
that restricting the meta-analysis to SNPs that are expression quantitative trait loci
eQTLs, p < 5\times10^{-8} with a gene expression trait in the populations\textsuperscript{17} for which we have
genome-wide gene expression data) allowed us to identify a highly significant SNP (after
Bonferroni correction, n = 4,686 tests), rs4073360 (p = 6.9x10^{-6}), that is a common (MAF > 5%) variant in all HapMap populations. Furthermore, the distribution of association p-values for the expression-associated SNPs (at a less stringent p < 10^{-4} with a gene expression trait)\textsuperscript{25,26} shows enrichment for low cytarabine association p-values.

A genic region that contains the highest-ranked SNP associations from the meta-analysis was identified on chromosome 5. Figure 2B illustrates a plot of the meta-analysis p-values for the genomic interval on chromosome 5 that contains the top signals; the plot also illustrates the degree of LD between the highest ranked SNP (rs7729269, p = 3.67 \times 10^{-7}, purple diamond) and surrounding SNPs in the region in CEU. The SNP, rs7729269, is a common variant (MAF of 0.27 in CEU, 0.06 in ASN, 0.16 in ASW, and 0.075 in YRI) in all populations and is located in an intron of \textit{MCC}. The \textit{MCC} gene harbors several of the highest ranked SNPs that are in strong LD (r^2 > 0.80), most of which (i.e., 10 of 12) show association with gene expression traits (p < 10^{-4}) as potential eQTLs. In particular, rs7729269 is a potential eQTL (in CEU) for the ectodysplasin A2 receptor (\textit{EDA2R}) gene. We found that higher gene expression of \textit{EDA2R} was significantly associated with cellular resistance to cytarabine in CEU (p=0.02) and in YRI (p=0.004).

The most significant LCL cytotoxicity SNP rs7729269 showed nominally significant associations (p < 0.05) with concordant direction of effect in YRI, ASW and ASN populations (Figure 3A). Each additional C allele results in lower AUC in the cytotoxicity assay and therefore confers sensitivity to cytarabine. Another SNP within \textit{MCC}, rs13171482, shows concordant direction in all 4 populations (Figure 3B).
**Evaluation of cytarabine-induced apoptosis**

Using the preclinical LCL model, we phenotyped 4 LCL panels (YRI1/2, ASW, CEU1/2, CHB) for caspase 3/7 activation following treatment with 40µM cytarabine. Upon evaluation of the top 37 cytotoxicity-associated SNPs (p < 1 x 10^{-5}) for their association with cytarabine-induced caspase 3/7 activation, 14 SNPs showed nominal association with apoptosis at a p < 0.05 (Table 1 bolded SNPs). Twelve of the 14 SNPs were within the *MCC* gene. Thus, the *MCC* SNPs are associated with greater cytarabine-induced cytotoxicity through activation of caspase 3/7.

**Evaluation of relationship of GWAS findings with intrinsic growth**

Previously, we demonstrated a strong association between the rate of cellular proliferation and sensitivity to cytarabine, as measured by AUC, within each population.

17,27 Nevertheless, we found no association (p>0.05) between any of the top 37 SNPs and intrinsic growth rate. Thus, it is unlikely that the observed associations of the 37 SNPs with cytarabine-induced cytotoxicity were mediated by the rate of basal proliferation. Furthermore, the expression of the *MCC* gene was not correlated with intrinsic growth rate (p = 0.11).

**Validation of SNPs from LCL model in clinical samples**

To determine the degree to which our cell-based approach identified patients responsive to cytarabine, we evaluated SNPs (n=37) identified in LCLs at p < 1 x 10^{-5} threshold (Table 1) in AML patients from a previously reported clinical trial. Thirty three SNPs were successfully genotyped and tested in the patient population. Among
these SNPs, 12 SNPs (located in the gene $MCC$) were in high LD (rs7729269, rs13189050, rs6594713, rs13181534, rs10053341, rs6870401, rs13171482, rs6887482, rs10061462, rs7714760, rs6594714, rs13174075). Located within the gene $SNYPR$, 5 SNPs (rs10510896, rs12638620, rs13068980, rs17068377, rs17068392) were in LD. Of the remaining SNPs, 4 SNPs on chromosome 3 (rs1533140, rs1567582, rs6550825 and rs6550826) were in LD with each other with $r^2 > 0.85$. The remaining 12 SNPs were not in LD with any other SNP and all were in Hardy-Weinberg equilibrium (HWE criteria is enforced within each population.)

We evaluated whether SNPs identified in the LCL model were associated with $in vitro$ chemo-sensitivity of leukemic blasts obtained at diagnosis ($n = 69$) or with treatment response in AML patients. Of the 33 SNPs tested that had an association with cytarabine cytotoxicity in LCLs, 10 SNPs within $MCC$ (as well as the intergenic eQTL, rs4956103) were found to be nominally significant ($p < 0.05$) with $in vitro$ cytarabine IC$_{50}$ in leukemic blasts (Table 1, Supplementary Table 2). These 10 SNPs also showed the same directionality for cellular sensitivity to drug as measured by LCL-cytotoxicity and LCL-apoptosis and for chemo-sensitivity of leukemic blasts from AML patients, as illustrated for rs7729269 and rs13171482 (in Figure 3C and 3D, Table 1, Supplementary Table 2).

Association of SNPs was also evaluated with clinical endpoints including minimal residual disease (MRD) on day 22, relapse-free survival (RFS), event-free survival (EFS), overall survival (OS), and treatment-related mortality (TRM). Of the 33 SNPs evaluated in all patients, rs2897047, a SNP near Iroquois-class homeodomain protein ($IRX2$) and associated with cellular sensitivity to cytarabine in LCLs ($p = 4.06 \times 10^{-6}$), was also
associated with day 22 MRD (p=0.021) and RFS (p = 0.043) (Figure 4). LCLs carrying TT genotype were resistant to cytarabine and patients with TT genotype did not respond as well to cytarabine as measured by MRD at day 22 (37.5% TT genotype patients had <0.1 MRD compared to 61.2% patients who had <0.1 MRD carrying CC and CT genotype). Consistent with cellular resistance to cytarabine in cells and higher MRD at day 22, those with a TT genotype had a poorer relapse-free survival (Figure 4C).

There were 5 SNPs (rs12036333, rs9883101, rs6550826, rs1533140, rs17202778) associated with TRM. Rare alleles were associated with greater sensitivity to cytarabine-induced cytotoxicity and with greater probability of TRM. Most TRM’s were due to infection secondary to chemotherapeutic immune suppression, one was due to veno-occlusive disease and another due to intracranial hemorrhage. The SNP rs12036333 was associated with LCL-cytotoxicity (p = 1.31 x 10^{-6}) with AA genotype associated with greater \textit{in vitro} cytarabine sensitivity and in AML patients, AA genotype was associated with inferior overall survival (p = 0.015) as well as greater treatment-related mortality (p = 0.0037) compared to those with GA/GG genotype (Figure 5).

Of the 33 SNPs tested that had an association with cytarabine-induced cytotoxicity in LCLs (p < 1 x 10^{-5}), 18 were associated (p<0.05) with cytarabine IC_{50} in leukemia cells, day 22 MRD, RFS, OS, or TRM (Supplemental Table 2). The number of such “clinically validated SNPs” was statistically significant (p=0.016; Figure 6) from a permutation analysis (see Methods).
Effect of Cytarabine treatment on MCC gene expression

There was no significant relationship between baseline MCC gene expression and sensitivity to cytarabine as measured by cytotoxicity or apoptosis (data not shown). We surmised that a potential mechanism for sensitivity to cytarabine by SNPs within MCC might involve variation in modulation of gene expression following drug treatment. To test whether MCC modulation occurred after cytarabine treatment, we treated 5 CEU and 5 YRI LCLs with vehicle alone or 10 µM cytarabine for 2, 6, 18, and 24 hours. We found all 10 cell lines demonstrated induction with variability observed in the degree and the time course ranging from 1.3 to 26.6 times MCC expression relative to control (Supplementary Figure 2). This induction was statistically significant at all time points tested (p=0.03 at 6 hr, 7.1x10^-10 at 18 hr, 7.1x10^-11 at 24 hr). Interestingly, the cell line (12044) with 26.6 fold induction was among the most sensitive and the cell line (12812) with 1.3 fold induction was among the most resistant to cytarabine. However, the relationship between fold change and cytarabine sensitivity failed to achieve significance in these 10 cell lines, perhaps due to small size. We also measured NFκB1 expression at the same time points and saw a modest expression decrease at 24 hours (p=0.059) (data not shown).

Using limma as implemented in the Bioconductor project, we re-analyzed two gene expression profiling datasets, one from a study of cytarabine-treated human diffuse large cell lymphoma cell lines versus untreated cells and another from a recently deposited data into Gene Expression Omnibus involving primary AML cells treated with cytarabine (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA174047). In the former, we found that MCC expression was significantly induced (p=0.029) by cytarabine treatment
(Supplementary Figure 3). In the latter, AML samples treated with cytarabine showed higher exon-level MCC expression (1.5 fold change, transcript isoform NM_001085377, Affymetrix Human Exon 1.0 ST Array probeset ID 2871377) relative to untreated samples (p=0.05) although probesets also annotated to a second transcript isoform for the gene (NM_002387) and showing concordant direction of effect did not attain nominal significance. These findings from two additional independent microarray datasets (GEO database under accession no. GSE5681 for human diffuse large cell lymphoma cell line and GSE40442 for AML samples) are consistent with the results of our experiments in LCLs showing induction of MCC.

Discussion

We performed a meta-analysis of the results of GWA studies for cytarabine-induced cytotoxicity in HapMap panels composed of 4 different populations (Caucasian, Asian, African American, African) that included a total of 523 LCLs. The top 37 SNPs (p < 10^{-5}) were also evaluated for cytarabine-induced apoptosis in LCLs and in a clinical cohort of AML patients. Fourteen of the 37 cytotoxicity SNPs were associated with apoptosis as measured by caspase 3/7 activation in the LCL model; of these, 12 are in LD within the introns of MCC. These same MCC variants were also significant in in vitro leukemic blasts (IC_{50}) with consistent directionality such that the allele most sensitive to cytarabine in the LCL model was also the allele associated with the greater sensitivity in patient samples. Although the role of MCC in cytarabine sensitivity in AML is not understood, we found that MCC expression was variably induced following cytarabine treatment in LCLs of CEU and YRI ancestry with fold increases ranging from 1.3 to 26.6.
In addition, an intergenic SNP (rs2897047) near IRX2 and associated with cellular sensitivity to cytarabine in LCLs was associated with day 22 minimal residual disease, and relapse-free survival in AML patients. The rare alleles of 5 SNPs were associated with cytarabine-induced cytotoxicity and greater probability of treatment-related mortality. Identifying patients upfront who are highly sensitive to cytarabine may have clinical utility and improve outcomes. In total, 18 SNPs were found to be associated with a clinical outcome in AML patients, which is significantly more than expected by chance.

Previously, we demonstrated that the cumulative incidence of relapse was significantly higher among those with high levels of MRD compared to patients with low levels of MRD (P<0.0001)\(^2\). In this study we have identified that carriers of the TT allele of rs2897047 (near IRX2) have greater incidence of relapse and more disease burden at day 22 as determined by MRD positivity consistent with our findings in LCLs that cells with TT genotype were more resistant to cytarabine. The SNP, rs2897047, is in linkage disequilibrium with rs6872448 (\(r^2 = 0.87\) in CEU) which is in a DNase hypersensitive site and likely to be a regulatory SNP for IRX2 based on ENCODE data in GM12891 and HeLaS3 cells\(^3\). Notably, three genes highly predictive of outcome in a recent Children’s Oncology Group study\(^3\) included IRX2. IRX2 expression was correlated with worse outcome in acute lymphoblastic leukemia (ALL) patients\(^3\).

Furthermore, of the 5 SNPs associated with cytarabine-induced cytotoxicity in LCLs and with treatment-related mortality in AML patients, 3 SNPs in high LD (\(r^2 > 0.80\)) – namely, rs9883101, rs6550826, and rs1533140 – flank the nuclear receptor subfamily 1, group D, member 2 (NR1D2) gene, which, in AML, shows significant (nominal p=0.018) evidence of non-silent mutations being elevated, on the basis of The
Cancer Genome Atlas data\textsuperscript{34}. Our preliminary analyses of data involving cytarabine treatment of AML cells show that cytarabine reduces the exon-level (p=0.015, transcript isoform NM\_005126, Affymetrix Human Exon 1.0 ST Array probeset ID 2614151) expression of \textit{NR1D2} in AML cells relative to untreated ones\textsuperscript{31}. Furthermore, this region has been found to harbor a cis eQTL, rs1567581, for \textit{NR1D2} in human monocytes (JK Pritchard’s eQTL resource)\textsuperscript{35} that is in complete LD (D' =1, \(r^2 = 0.132\)) with rs6550826 and has nominal association with cytarabine-induced cytotoxicity in LCLs (p=0.002) and concordant direction of effect in all population panels examined here. The impact of this locus on response to cytarabine thus merits further investigation.

Our data related to the MCC SNPs demonstrate that: 1) Genetic variants within \textit{MCC} are associated with cytarabine-induced cytotoxicity and apoptosis in LCLs; 2) \textit{MCC} variants are also associated with \textit{in vitro} cytarabine sensitivity of leukemic blasts from AML patients. 3) Several SNPs within \textit{MCC} (rs7729269, rs6594713, rs13181534, rs6870401, rs6887482, rs10061462, rs6594714, rs13174075) are potential eQTLs for \textit{EDA2R}, a gene shown to be a direct p53 target\textsuperscript{36} and involved in p53-mediated apoptosis\textsuperscript{37}. \textit{EDA2R} (also known as XEDAR) is a member of the TNFR superfamily that interacts with TRAFs (TNFR-associated factors) and activates the nuclear factor-\(\kappa\)B (NF-\(\kappa\)B) signaling, and consequently promotes cell proliferation. 4) Upon treatment of LCLs with cytarabine, there is an increase in \textit{MCC} that is variable within the CEU and YRI cell lines.

Because of the clinical importance of this drug, there have been numerous studies to uncover the biological mechanisms of resistance. Evidence suggests that cytarabine activates NF-\(\kappa\)B, a nuclear transcription factor, in myeloid cells\textsuperscript{38}. Activation of NF\(\kappa\)B may be accompanied by the acquisition of cytarabine resistance; the activation is
assumed to induce transcription of genes that function in a feedback and block apoptosis. Constitutive activation of NFκB is found in human leukemia stem cells but not in normal hematopoietic stem cells. Interestingly, the tumor suppressor gene MCC has been shown to be a transcriptional regulator of NFκB pathway in colorectal cells and HEK293 cells. Knockdown of MCC results in the accumulation of the inhibitor of kappa B alpha (IkBα) protein, encoded by NFKBIA, a first-response gene specifically and rapidly regulated by NFκB pathway activation. Indeed, MCC has been shown to modulate NFκB pathway signaling indirectly in colorectal cells. One feasible explanation for the role of MCC is that greater induction of this gene by cytarabine associates with suppression of NFκB and ultimately greater sensitivity to cytarabine.

GeneCard queries for the relative expression levels of MCC in a variety of tissues identified MCC in malignant cells of the bone marrow, spleen and whole blood. Of note, we also found that increased cytosine modification at the CpG cg01272202 was correlated with both lower expression of MCC and reduced cytarabine sensitivity in the CEU (p=0.005) and YRI (p=0.076) cell lines (W. Zhang and M. E. Dolan, unpublished data). Further studies of the contribution of MCC genetic variation and expression to cytarabine sensitivity are warranted.

We evaluated a set of SNPs in known candidate genes and identified nominally significant (p < 0.05) SNPs in CDA, SLC29A1 and NT5C2. For example, SNPs in the 5'-untranslated region of NT5C2 were associated with NT5C2 expression and cytarabine sensitivity in the HapMap cell lines and with NT5C2 mRNA expression and cytarabine sensitivity in diagnostic leukemic blasts from pediatric patients with AML. The NT5C2 SNP rs4917384 has been shown to be significantly associated with induction 1 response.
(measured as day 22 MRD) in AML patients. In AML patients, higher CDA levels have been associated with disease recurrence and lower CDA levels with longer duration of remission. Identifying (nominally) significant SNPs within candidate genes provides some level of confidence in the GWAS results.

It is plausible that some pharmacogenetic effects may be dose-specific. However, our study did not have an adequate number of subjects to provide statistically robust results for dose-specific pharmacogenetic effects. Previous trials use different doses of cytarabine and combine cytarabine with different agents. Thus, we chose to perform an arm-stratified statistical analysis to identify pharmacogenetic effects that are not strongly impacted by dose. We believe that these types of pharmacogenetic effects have greater potential to be confirmed in independent cohorts and thus be translated to improve clinical practice. Finally, as previously described by, the two arms differ very little in terms of any of the clinical outcome, including those considered in this study. For example, there were 4 TRMs in each arm (low-dose and high-dose ara-C).

Recent studies from a number of groups using pre-clinical cell-based models have taken top GWAS findings identified in LCLs and validated them in prospective clinical trials. The current study further demonstrates the utility of the model system and provides an overall measure of the significance of the findings from its use in predicting response in AML patients treated with cytarabine.
Acknowledgements. This work was supported by a Specialized Center of Research Grant from the Leukemia and Lymphoma Society (MED), NIH/NIGMS U01 GM61393 (MED, NJC), Department of Defense Breast Cancer Research Program grant BC087674 (MW), NIH/NCI Cancer Biology Training grant T32CA009594 (HEW), NIH/NIGMS Clinical Therapeutics Training Grant 5T32GM007019 (CMH), NIH R01CA132946 (JL), Cancer Research Foundation of the University of Chicago Comprehensive Cancer Center (MED) and R21HG006367 (to WZ, MED). The authors are grateful to the PAAR LCL core for maintaining and distributing lymphoblastoid cell lines as well as Dr. Stephanie Huang for helpful discussions.

Authorship and Conflict of Interest Statements

Designed research (ERG, JL, ALS, MED), collected data, analyzed and interpreted preclinical data (ERG, ALS, SSW, MW, IH, LG, CMH, WZ), contributed analytical tools (ERG, NJC), designed, directed, or collected data for the AML02 clinical trial (SP, JER, RCR, SR, DC, KRC), genotyped clinical samples (JL, AKM), performed statistical analysis (ERG, SP, HEW, XC, HKI), wrote the manuscript (ERG, JL, ALS, MED).

There are no conflicts of interest.
References

Table 1. Summary of top signals from a meta-analysis of cellular sensitivity to cytarabine (AUC). These SNPs were evaluated in the patient population.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Allele 1</th>
<th>Allele 2</th>
<th>Chrom</th>
<th>Position</th>
<th>Weight $^2$</th>
<th>P-value</th>
<th>Direction of Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7729269</td>
<td>t</td>
<td>c</td>
<td>5</td>
<td>112748293</td>
<td>502</td>
<td>3.67E-07</td>
<td>++++++</td>
</tr>
<tr>
<td>rs13189050</td>
<td>t</td>
<td>c</td>
<td>5</td>
<td>112734410</td>
<td>502</td>
<td>4.11E-07</td>
<td>++++++</td>
</tr>
<tr>
<td>rs6594713</td>
<td>a</td>
<td>c</td>
<td>5</td>
<td>112743273</td>
<td>502</td>
<td>6.06E-07</td>
<td>++++++</td>
</tr>
<tr>
<td>rs13181534</td>
<td>t</td>
<td>c</td>
<td>5</td>
<td>112752540</td>
<td>502</td>
<td>6.06E-07</td>
<td>++++++</td>
</tr>
<tr>
<td>rs17068377</td>
<td>t</td>
<td>g</td>
<td>3</td>
<td>63327579</td>
<td>254</td>
<td>6.52E-07</td>
<td>---????</td>
</tr>
<tr>
<td>rs12638620</td>
<td>c</td>
<td>g</td>
<td>3</td>
<td>63329590</td>
<td>253</td>
<td>7.17E-07</td>
<td>+++??</td>
</tr>
<tr>
<td>rs17068392</td>
<td>t</td>
<td>c</td>
<td>3</td>
<td>63331024</td>
<td>253</td>
<td>7.17E-07</td>
<td>+++??</td>
</tr>
<tr>
<td>rs10510896</td>
<td>a</td>
<td>g</td>
<td>3</td>
<td>63330132</td>
<td>254</td>
<td>7.60E-07</td>
<td>---????</td>
</tr>
<tr>
<td>rs13068980</td>
<td>a</td>
<td>g</td>
<td>3</td>
<td>63325509</td>
<td>253</td>
<td>7.72E-07</td>
<td>+++??</td>
</tr>
<tr>
<td>rs10053341</td>
<td>t</td>
<td>g</td>
<td>5</td>
<td>112811906</td>
<td>511</td>
<td>8.18E-07</td>
<td>------</td>
</tr>
<tr>
<td>rs6870401</td>
<td>c</td>
<td>g</td>
<td>5</td>
<td>112766263</td>
<td>510</td>
<td>8.74E-07</td>
<td>++++++</td>
</tr>
<tr>
<td>rs13171482</td>
<td>a</td>
<td>g</td>
<td>5</td>
<td>112729865</td>
<td>509</td>
<td>9.29E-07</td>
<td>------</td>
</tr>
<tr>
<td>rs6887482</td>
<td>a</td>
<td>c</td>
<td>5</td>
<td>112769332</td>
<td>501</td>
<td>1.02E-06</td>
<td>++++++</td>
</tr>
<tr>
<td>rs12036333</td>
<td>a</td>
<td>g</td>
<td>1</td>
<td>237585689</td>
<td>175</td>
<td>1.31E-06</td>
<td>--????</td>
</tr>
<tr>
<td>rs10061462</td>
<td>a</td>
<td>t</td>
<td>5</td>
<td>112754104</td>
<td>502</td>
<td>1.67E-06</td>
<td>++++++</td>
</tr>
<tr>
<td>rs6550825</td>
<td>a</td>
<td>g</td>
<td>3</td>
<td>24009973</td>
<td>508</td>
<td>2.29E-06</td>
<td>++++++</td>
</tr>
<tr>
<td>rs4956103</td>
<td>c</td>
<td>g</td>
<td>4</td>
<td>108619495</td>
<td>510</td>
<td>2.91E-06</td>
<td>++++++</td>
</tr>
<tr>
<td>rs10758713</td>
<td>a</td>
<td>g</td>
<td>9</td>
<td>5862949</td>
<td>253</td>
<td>3.25E-06</td>
<td>---????</td>
</tr>
<tr>
<td>rs7714760</td>
<td>t</td>
<td>c</td>
<td>5</td>
<td>112763130</td>
<td>487</td>
<td>3.45E-06</td>
<td>++++++</td>
</tr>
<tr>
<td>rs1567582</td>
<td>a</td>
<td>c</td>
<td>3</td>
<td>24011171</td>
<td>508</td>
<td>3.68E-06</td>
<td>------</td>
</tr>
<tr>
<td>rs6594714</td>
<td>c</td>
<td>g</td>
<td>5</td>
<td>112766519</td>
<td>509</td>
<td>3.91E-06</td>
<td>++++++</td>
</tr>
<tr>
<td>rs2897047</td>
<td>a</td>
<td>g</td>
<td>5</td>
<td>2693123</td>
<td>492</td>
<td>4.06E-06</td>
<td>++++++</td>
</tr>
<tr>
<td>rs7122539</td>
<td>a</td>
<td>g</td>
<td>11</td>
<td>66419307</td>
<td>489</td>
<td>4.19E-06</td>
<td>------</td>
</tr>
<tr>
<td>rs9883101</td>
<td>a</td>
<td>c</td>
<td>3</td>
<td>24019404</td>
<td>502</td>
<td>4.26E-06</td>
<td>++++++</td>
</tr>
<tr>
<td>rs16873946</td>
<td>a</td>
<td>g</td>
<td>4</td>
<td>23239147</td>
<td>90</td>
<td>4.44E-06</td>
<td>+??????</td>
</tr>
<tr>
<td>rs8037890</td>
<td>t</td>
<td>g</td>
<td>15</td>
<td>34953560</td>
<td>162</td>
<td>4.49E-06</td>
<td>????+</td>
</tr>
<tr>
<td>rs17808412</td>
<td>c</td>
<td>g</td>
<td>3</td>
<td>108622000</td>
<td>164</td>
<td>4.63E-06</td>
<td>?--???</td>
</tr>
<tr>
<td>rs1533140</td>
<td>a</td>
<td>g</td>
<td>3</td>
<td>24014384</td>
<td>508</td>
<td>4.81E-06</td>
<td>++++++</td>
</tr>
<tr>
<td>rs6078860</td>
<td>t</td>
<td>c</td>
<td>20</td>
<td>12914738</td>
<td>164</td>
<td>4.99E-06</td>
<td>?--???</td>
</tr>
<tr>
<td>rs6550826</td>
<td>c</td>
<td>g</td>
<td>3</td>
<td>24009992</td>
<td>508</td>
<td>5.40E-06</td>
<td>++++++</td>
</tr>
<tr>
<td>rs12498793</td>
<td>t</td>
<td>c</td>
<td>4</td>
<td>108617636</td>
<td>509</td>
<td>5.49E-06</td>
<td>++++++</td>
</tr>
<tr>
<td>rs13174075</td>
<td>a</td>
<td>t</td>
<td>5</td>
<td>112823251</td>
<td>510</td>
<td>6.18E-06</td>
<td>++++++</td>
</tr>
<tr>
<td>rs4073360</td>
<td>a</td>
<td>g</td>
<td>12</td>
<td>37161419</td>
<td>255</td>
<td>6.92E-06</td>
<td>--???</td>
</tr>
<tr>
<td>rs6883547</td>
<td>a</td>
<td>g</td>
<td>5</td>
<td>112944752</td>
<td>501</td>
<td>6.93E-06</td>
<td>------</td>
</tr>
<tr>
<td>rs6714052</td>
<td>t</td>
<td>c</td>
<td>2</td>
<td>6691777</td>
<td>163</td>
<td>7.30E-06</td>
<td>?--???</td>
</tr>
<tr>
<td>rs17202778</td>
<td>t</td>
<td>c</td>
<td>2</td>
<td>207943031</td>
<td>254</td>
<td>8.62E-06</td>
<td>+++??</td>
</tr>
<tr>
<td>rs12712936</td>
<td>a</td>
<td>g</td>
<td>2</td>
<td>45244339</td>
<td>510</td>
<td>8.93E-06</td>
<td>------</td>
</tr>
</tbody>
</table>

1 Chrom refers to chromosome of host SNP. 2 Weight refers to the number of samples analyzed. 3 Order of populations: ASN, CEU1/2, CEU3, ASW, YRI1/2, YRI3; the "--"
means the SNP minor allele is absent in the corresponding population or the SNP has too low a MAF in that population for GWAS or the SNP did not meet a quality control threshold (e.g., imputation $R^2$) for GWAS. Bold indicates SNPs associated with IC$_{50}$ in LCLs (p-value shown) as well apoptosis.
Figure Legends

Figure 1. Cellular sensitivity to cytarabine in 523 HapMap LCLs. LCLs from world populations (ASN, ASW, CEU, YRI) were phenotyped for the growth inhibitory effects of cytarabine. The mean (±SD) log$_2$ AUC was 10.81 (±0.30 μM) for ASN, 10.78 (±0.27 μM) for ASW, 10.83 (±0.30 μM) for CEU, and 10.78 (±0.29 μM) for YRI.

Figure 2. Meta-analysis of GWAS in CEU, YRI, ASW and ASN populations. (A) The Manhattan plot shows the meta-analysis results on GWAS of cellular susceptibility to cytarabine in 4 world populations (n=523 individual samples). The highest ranked SNP was rs7729269 (p = 3.67 × 10^{-7}) in the gene MCC. (B) The plot shows p-values from the meta-analysis of GWAS in CEU, YRI, ASW and ASN populations for the top SNP associations with cytarabine log2 AUC in a specific region on chromosome 5. The colors, as indicated in the legend, denote the extent of LD between the SNPs; for this purpose we use only the CEU reference population. The bottom panel illustrates the chromosomal region and genes that these SNPs fall on. The figure was made in LocusZoom.

Figure 3. Association of SNPs within MCC with cytarabine sensitivity in LCLs and in leukemic blasts. MCC SNPs plotted against log2 AUC in lymphoblastoid cell lines for (A) rs7729269 and (B) rs13171482, and MCC SNPs plotted against IC$_{50}$ (dose required to inhibit growth of 50% of cells) in leukemic blasts for (C) rs7729269 and (D) rs13171482.
Figure 4. SNP rs2897047 association with log2 AUC in LCLs and association with minimal residual disease and RFS in all AML patients. Association of SNP rs2897047 with (A) log2 AUC for cellular sensitivity to cytarabine in cell-based model (meta $p = 4.06 \times 10^{-6}$) illustrated in CEU, YRI, ASN; (B) minimal residual disease at day 22 in all AML patients ($p = 0.0218$), the numbers in each box represent MRD levels <0.1, 0.1-1 or $\geq 1$, number in parentheses indicate fraction of patients within each MRD category; (C) relapse-free survival ($p = 0.043$) in all AML patients.

Figure 5. SNP rs12036333 association with log2 AUC in LCLs and association with overall survival and treatment-related mortality. Association of SNP rs12036333 (AA versus GA/GG) with (A) log2 AUC for cellular sensitivity to cytarabine in cell-based model in all populations ($p = 0.007$); (B) overall survival ($p = 0.0146$) and (C) treatment-related mortality in white AML patients ($p = 0.0037$).

Figure 6. Top SNPs identified in LCLs are enriched for top associations with clinical phenotypes. Of the 33 SNPs tested that had an association with cytarabine-induced cytotoxicity in LCLs ($p < 1 \times 10^{-5}$), 18 were associated ($p < 0.05$) with cytarabine IC$_{50}$ in leukemia cells, day 22 MRD, OS, or TRM. The number of “clinically validated SNPs” is highly significant ($p = 0.016$) given the correlation structure of the genotypes evaluated and that of the phenotypes examined.
Figure 1
Figure 2

A

-\log_{10} (p-value)

Position on chrom 5 (Mb)

B

CEU

-\log_{10} (p-value)

Position on chrom 5 (Mb)

rs7729269
Figure 4

rs2897047 (p=0.043)
Figure 5

A. Box plots showing Cytarabine Log_{2}AUC for different genotypes (AA, GA, GG) for CEU, YRI, ASW, and ASN populations. AA vs. GA+GG, P=0.0146.

B. Kaplan-Meier curves for the probability of overall survival with AA vs. GA+GG P=0.0146. AA:9 vs. GG:91.

C. Kaplan-Meier curves for the probability of treatment-related mortality with AA vs. GA+GG P=0.0037. AA:6 vs. GG:91.
Figure 6
Comprehensive genetic analysis of cytarabine sensitivity in a cell-based model identifies polymorphisms associated with outcome in AML patients