ORIGINAL ARTICLE

Impact of polymorphisms in drug pathway genes on disease-free survival in adults with acute myeloid leukemia

Sook Wah Yee¹, Joel A Mefford², Natasha Singh³, Mary-Elizabeth Percival³, Adrian Stecula¹, Kuo Yang¹, John S Witte², Atsushi Takahashi⁴, Michiaki Kubo⁴, Koichi Matsuda⁵, Kathleen M Giacomini^{1,6,7} and Charalambos Andreadis^{3,6,7}

Acute myeloid leukemia (AML) is a clinically heterogeneous disease, with a 5-year disease-free survival (DFS) ranging from under 10% to over 70% for distinct groups of patients. At our institution, cytarabine, etoposide and busulfan are used in first or second remission patients treated with a two-step approach to autologous stem cell transplantation (ASCT). In this study, we tested the hypothesis that polymorphisms in the pharmacokinetic and pharmacodynamic pathway genes of these drugs are associated with DFS in AML patients. A total of 1659 variants in 42 genes were analyzed for their association with DFS using a Cox-proportional hazards model. One hundred and fifty-four genetically European patients were used for the primary analysis. An intronic single nucleotide polymorphism (SNP) in *ABCC3* (rs4148405) was associated with a significantly shorter DFS (hazard ratios (HR) = 3.2, $P = 5.6 \times 10^{-6}$) in our primary cohort. In addition, a SNP in the *GSTM1-GSTM5* locus, rs3754446, was significantly associated with a shorter DFS in all patients (HR = 1.8, $P = 0.001$ for 154 European ancestry; HR = 1.7, $P = 0.028$ for 125 non-European patients). Thus, for the first time, genetic variants in drug pathway genes are shown to be associated with DFS in AML patients treated with chemotherapy-based autologous ASCT.

Journal of Human Genetics (2013) 58, 353–361; doi:10.1038/jhg.2013.38; published online 16 May 2013

Keywords: acute myeloid leukemia; autologous stem cell transplant; busulfan; cytarabine; etoposide; pharmacogenomics

INTRODUCTION

Adult acute myeloid leukemia (AML) is a hematologic malignancy with widely heterogeneous clinical outcomes. New treatments for AML are increasingly being tested in clinical trials of patients with specific tumor cell mutations.^{1,2} Although there have been substantial improvements in the number of patients who achieve complete remission, the choice of induction and post-remission therapy for adult AML is still based on the 'one size fits all' principle. Most regimens incorporate antimetabolites (for example, cytarabine, fludarabine), topoisomerase II inhibitors (for example, etoposide, daunorubicin, idarubicin, mitoxantrone) and alkylating agents (for example, busulfan, cyclophosphamide) for the treatment of AML. Prognostic factors for treatment response include age, prior exposure to chemotherapy, cytogenetic markers and expression profiles, and appearance of specific genetic mutations in tumor tissue, such as mutation and translocation of particular genes (for example, *FLT3*, *NPM1*).^{3,4} However, these prognostic factors do

not adequately capture the wide diversity of clinical outcomes in this disease. The percent of adults with AML who can survive ≥ 3 years and may be cured is $\sim 5\text{--}70\%$.^{5,6}

One possible explanation for the difference in response to AML treatment is germline genetic variation. Although the pharmacogenomics of AML drug response is an active area of research, there remain large challenges. These include the: (i) poor availability of uniform, well-collected and well-defined drug response phenotype information; (ii) lack of widely available germline DNA not contaminated with tumor cells (myeloblasts); (iii) limited availability of panels of genotype data in large patient cohorts; and (iv) inability to validate findings in replication studies. In this study, we overcome many of these challenges. In particular, to our knowledge, this is the first study that involved the analysis of large numbers of genetic polymorphisms in a cohort of AML patients treated with high-dose chemotherapy followed by autologous stem cell transplantation (ASCT). We present results from testing the association between

¹Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, USA; ²Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, USA; ³Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, USA; ⁴Center for Genomic Medicine, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan; ⁵Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan and ⁶Institute for Human Genetics, University of California, San Francisco, CA, USA

⁷Co-senior authors.

Correspondence: Dr C Andreadis, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA.

E-mail: candreadis@medicine.ucsf.edu

Received 18 March 2013; accepted 11 April 2013; published online 16 May 2013

germline variants in drug pathway and other genes with disease-free survival (DFS) in adult AML patients. We identified new associations between AML DFS and polymorphisms in several drug pathway genes for cytarabine, etoposide and busulfan, and also replicated single nucleotide polymorphisms (SNPs) previously reported to be associated with the AML response phenotypes.

MATERIALS AND METHODS

Clinical protocol, study criteria and patient cohorts

The population for the current study consists of AML patients who were enrolled in University of California, San Francisco (UCSF) study protocols 9203 or 9303 between 1988 and 2010. The study protocols and patient selection criteria have been previously described.^{7–9} This treatment protocol was used in patients with low- and standard-risk *de novo* AML, including acute promyelocytic leukemia in first or second complete remission. It was also used in a small number of patients with high risk AML (that is, with secondary AML) if allogeneic stem cell transplantation was not an option for the patient (for example, unavailable donor).⁷ In step 1, patients were treated with consolidation chemotherapy including cytarabine 2000 mg m⁻² (intravenously) twice daily for 4 days concurrently, with etoposide 40 mg kg⁻¹ by intravenous infusion over the 4 days. During the recovery period from chemotherapy, peripheral blood stem cells were collected under granulocyte colony-stimulating factor stimulation. In step 2, patients underwent ASCT, which involved the preparative regimen of busulfan (total dose 16 mg kg⁻¹ orally or 12.8 mg kg⁻¹ intravenously, over 16 doses in 4 days), followed by etoposide 60 mg kg⁻¹ (intravenous bolus) and reinfusion of blood or marrow stem cells. Patients had to be in complete remission for at least 30 days before step 2 (Figure 1). Complete remission was defined as normal bone marrow morphology with <5% blasts, resolution of previously abnormal cytogenetics and no evidence of extramedullary leukemia. In addition, patients must meet criteria for neutrophil and platelet counts, liver and kidney function.^{7–9} Detailed procedures of patient enrollment, diagnosis, data collection and follow-up have been previously described.^{7–9} Briefly, patients were actively followed-up in the beginning within 6 months of diagnosis, with subsequent annual follow-up by clinic visits. UCSF electronic medical records, the UCSF Blood and Bone Marrow Transplant Clinic database and patients' medical charts were abstracted to determine patients' remission status. The UCSF Committee on Human Research approved the research protocol (institutional review board number 10-00649).

DNA isolation and genotyping

DNA was isolated from peripheral blood stem cells, which were collected during the recovery from step 1 consolidation chemotherapy. As noted in the above section, patients were in complete remission before consolidation chemotherapy, and hence the samples utilized in this step contained <5% leukemic cells. DNA was isolated at the UCSF DNA Banking and Extraction Services Lab. The lab followed standard DNA extraction protocol described in the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA). The DNA was then quantified using Picogreen (Life Technologies, Grand Island, NY, USA) and normalized to 50 ng µl⁻¹. For each sample, we genotyped 250 ng of DNA. The Illumina HumanOmniExpress v1.0 Beadchip was used, following the manufacturer's protocols, at the Center for Genomic Medicine, RIKEN, Yokohama, Japan. For quality control of the genotyping, we included one HapMap trio and three duplicates of the DNA samples from the AML patients. A total of 328 distinct DNA samples from patients were genotyped, along with three duplicates and three HapMap samples (trio).

Patients' ancestral origin

The genetic ancestral origin of patients was determined using principal component analysis implemented in Eigenstrat.¹⁰ Genotype information on our 328 AML patients was analyzed in conjunction with SNP data from the HapMap project, which consists of Europeans (CEU and TSI), Asians (JPT and CHB), Africans (TSI and ASW) and Mexican (MXL). From these analyses we were able to distinguish 154 patients of European ancestry on the basis of their close clustering with the European HapMap samples.

Pathway, gene and SNP selection

A total of 42 genes were selected for analysis based on the following criteria: (i) genes in the pharmacokinetic and/or pharmacodynamics pathway of the drugs administered (cytarabine, etoposide and busulfan);^{11–15} (ii) genes described in literature as having significant associations with drug cytotoxicity in lymphoblastoid cell lines (LCL);^{16,17} (iii) genes involved in DNA mismatch repair;¹⁸ and (iv) genes found previously to be associated with the AML response phenotype¹⁹ (see Supplementary Table 1, Figure 1). After filtering the SNPs with low call rates (<90%) and SNPs with low minor allele frequencies (MAF; MAF<1%) in the 154 European ancestry patients, we selected the SNPs in the candidate genes and within 25 000 bp upstream and downstream flanking regions.

Statistical analysis of the associations

The primary analysis was to estimate the association between SNPs in the selected candidate genes and DFS in 154 patients of European ancestry. A Cox-proportional hazard model was used to estimate the hazard ratios (HR) and 95% confidence limits for the effect of genotype on DFS. An additive coding of genotypes was used in all analyses. The genetic effect estimates were adjusted for levels of a clinical risk score (see Table 1). The SNP associations with $P<0.01$ were also tested in the 125 non-European patients using a Cox-proportional hazard model. In light of the heterogeneous ethnicity of the samples, the genetic effect estimates were adjusted for the first 10 principal components calculated from the genome-wide association study data and the clinical risk score. The Cox-proportional hazards function from the R-project (version 2.15.1) was used. We used 3×10^{-5} (0.05/1659) as the significance level after Bonferroni correction for multiple testing.

Fine mapping of associations *via* imputation

In order to further clarify the association signals, we performed imputation on genes with $P<0.005$ in the primary analysis. For eight genes (*ABCC3*, *DCK*, *GSTM1*, *GSTT1*, *MSH3*, *RRM1*, *SLC22A12*, and *SLC28A3*), the genotypes at polymorphic sites known from the 1000 Genomes Project but not observed on the Illumina HumanOmniExpress v1.0 Beadchip were imputed using IMPUTE version 2 (version 2.3.0 for Mac OS X, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html). The reference panel for the imputation was the 1000 Genomes Phase I-integrated variant set referenced to NCBI build b37 (March 2012 release, retrieved January 20 2013 from http://mathgen.stats.ox.ac.uk/impute/impute_v2.html#reference). Imputed variants with imputation quality scores <0.3 or MAF <0.01 were excluded. The remaining imputed SNPs were each used in Cox-proportional hazards models to predict DFS, as were the genotyped SNPs.

Functional studies

The potential functional effects of the SNPs with $P<0.01$ associated with DFS in European population were examined using the following steps:

- (1) All tag-SNPs in linkage disequilibrium to the SNPs with $P<0.01$ in our primary analysis were identified using the Proxy Search in the Broad Institute SNAP (SNP Annotation and Proxy Search; version 2.2), <http://www.broadinstitute.org/mpg/snap/ldsearch.php>. The search options used in this step were: SNP data set = 1000 Genomes Pilot 1 in CEU population panel, r^2 threshold = 0.8 and distance limit = 500 kbp.
- (2) Potential regulatory functions were identified by searching the following databases: Regulome database²⁰ and eQTL browser (<http://eqtl.uchicago.edu>, <http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>).
- (3) Literature searches were used to identify previous reports of associations with drug response phenotypes of any of the SNPs associated with AML response in this study.

RESULTS

This study investigates the potential associations between variants in 42 candidate genes and DFS in adult AML patients treated with a two-step approach to ASCT. High-dose cytarabine, etoposide and busulfan were used in this treatment approach. Table 1 describes

Figure 1 Schematic of workflow applied to determine the association of genetic variations in 42 candidate genes with disease-free survival (DFS) in acute myeloid leukemia (AML) patients treated with a two-step treatment protocol before autologous stem cell transplantation (ASCT). The workflow begins with phenotype data collection and genotyping of DNA samples using a genome-wide SNP array. Candidate genes were selected on the basis of their roles in the: drugs' pharmacokinetic/pharmacodynamics pathway; DNA mismatch repair mechanism; association with the drug cytotoxicity in lymphoblastic cell lines previously identified in literatures; and pharmacogenomics studies of AML drug response. After association of each SNPs with DFS in the 154 AML patients of European ancestry, the SNPs with $P < 0.01$ and their tag-SNPs were examined for their potential functional roles using databases to identify eQTL SNPs and predicted regulatory elements, such as binding sites of transcription factors and for their replications of previous studies. The SNPs with $P < 0.01$ were also examined for their associations with DFS in 125 non-European ancestry. Finally, imputation was performed to identify other SNPs with stronger associations with DFS in European ancestry.

the demographic characteristics of this cohort, which consisted of 154 patients of European ancestry and 125 patients of non-European ancestry (African, Asian and Mexican). In the DFS analysis, 55 patients of European ancestry (35.7%) and 47 of non-European ancestry (37.6%) relapsed during the observation period, which extended from 1988 to 2010.

We applied quality control criteria to the SNP data, with a genotype call rate of 0.99 and minor allele frequency ≥ 0.01 . After filtering, a total of 42 genes covered by 1659 SNPs were included in this association analysis. Furthermore, none of the DNA samples showed chromosomal abnormality by GenomeStudio (Illumina). Results in Figure 2 and Table 2 showed that among the 40 SNPs

with $P < 0.01$, the SNPs in the pharmacokinetic/pharmacodynamic pathway genes have stronger associations compared with SNPs not in the drug pathway. After Bonferroni correction for multiple testing (1659 tests), a SNP in the first intron of *ABCC3* (rs4148405) was significantly associated with DFS, with the minor allele (G) associated with shorter time to relapse (P -adjusted = 9.5×10^{-6} , Figure 3a). Although other SNPs did not reach significance after Bonferroni correction, overall there were 23 SNPs associated with DFS at

Table 1 Demographic characteristics of patients undergoing autologous stem cells transplantation for AML from 1988 to 2010

Variables	Patients of European ancestry (N = 154)	Patients of non-European ancestry (N = 125)
Sex, N (%)		
Male	76 (49.3%)	67 (53.6%)
Female	78 (50.6%)	58 (46.4%)
Age (years)		
Median (s.d., range)	47.0 ^a (13.0, 18–72)	40.5 ^b (12.8, 17–68)
Age at diagnosis	47.0 (13.0, 19–72)	41.0 (12.8, 19–69)
Age at autologous bone marrow transplantation		
Year of transplantation		
1988–1995	36 (23.4%)	38 (30.4%)
1996–2000	41 (26.6%)	31 (24.8%)
2001–2010	77 (50.0%)	56 (44.8%)
De novo/secondary AML, N (%)		
De novo	147 (95.5%)	110 (88.0%)
Secondary	7 (4.5%)	15 (12.0%)
Risk, N (%) ^c		
Acute promyeloid leukemia	15 (9.7%)	12 (9.6%)
Low	18 (11.7%)	18 (14.4%)
Standard	114 (74.0%)	80 (64.0%)
High	7 (4.5%)	15 (12.0%)
DFS (months)		
Median (s.d.)	21.4 (43.9)	16.7 (44.8)

Abbreviations: AML, acute myeloid leukemia; DFS, disease-free survival.

^aN = 152.

^bN = 124.

^cDefined by the Cancer and Leukemia Group B (CALGB) criteria.

$P < 0.005$ (Table 2). These 23 SNPs are in or within 25 000 bp of eight genes: *SLC28A3*, *DCK*, *RRM1*, *GSTM1*, *ABCC3*, *MSH3*, *GSTT1*, or *SLC22A12*. The majority of the minor alleles were associated with poor outcome (shorter DFS). Kaplan–Meier estimate plots of DFS are shown in Figures 3a–d for four of the top SNPs with MAF $\geq 3\%$. Other SNPs in the genes of the cytarabine pathway including *NT5C2* and *RRM2B* were also associated with DFS but with weaker P -values ($P < 0.01$, Table 2). Interestingly, SNPs in three out of eight selected genes (*SLC25A37*, *WNT5B* and *P2RX1*), for which expression levels have previously been correlated with either etoposide or cytarabine IC₅₀ values in LCL, showed significant but weaker association ($P < 0.01$) compared with genes in the drug pathways. Next, we examined the 40 SNPs in patients of non-European ancestry. Only one SNP, in *GSTM1-GSTM5* locus (rs3754446), was significantly associated with DFS in patients of non-European ancestry (HR = 1.7, $P = 0.028$). Overall, in the entire cohort the minor allele of the SNP (rs3754446) was significantly associated with shorter DFS (HR = 1.7, $P = 0.00027$).

Imputations of the eight candidate genes were performed to determine whether other SNPs in the regions have stronger association with DFS. The results showed that there are 234 imputed SNPs with $P < 0.01$ (MAF $\geq 1\%$), and among these there are 93 SNPs with improved P -values compared with the genotyped SNPs (Figure 4). Several SNPs in deoxycytidine kinase (*DCK*) have significant P -values $< 1.0 \times 10^{-4}$ (Figure 4), with MAF 5–10%. Although none of these 93 SNPs are in exonic regions, on examination in the eQTL browser, Regulome database and GTEx eQTL browser, we determined that SNPs in *GSTM1* (rs929166, rs11101989) and *MSH3* (6151896) are associated with their respective gene expression levels in liver²¹ or lymphoblastoid cells²² (data not shown). In the *ABCC3* and *SLC28A3* regions, imputation analysis did not identify other more significant SNPs in addition to the most significant genotyped SNPs, rs4148405 and rs11140500, respectively (Figure 4).

Using *in silico* analysis, we determined whether the 40 SNPs and their tag-SNPs were in known or predicted functional regions of the genome. Several of the SNPs were in DNA regions predicted to have binding sites for transcription factors. Some of these regions appear in

Figure 2 Plot showing the significance ($-\log_{10}$ of the P -value) of associations of the 1659 SNPs with DFS in 154 acute myeloid leukemia (AML) patients of European ancestries. Only SNPs with minor allele frequencies (MAF) of $\geq 1\%$ in the selected 42 candidate genes are shown. Each dot represents a SNP. SNPs above the black dotted line are SNPs with $P < 0.01$, and the SNP above the red dotted line reached a P -value, which was significant after correction for multiple testing ($P < 3 \times 10^{-5}$).

Table 2 Significant SNPs ($P < 0.01$) associated with DFS in 154 AML patients of European ancestry

SNP	Chromo- some	Gene	Feature	5' flanking gene/ 3' Flanking gene	Risk allele	MAF	P (unadjusted)	HR	(95% CI) of the HR	Genotype counts	Minor allele	Major allele	Reason for gene selection
rs4148405	17	ABCC3	Intron	CACNA1G/ANKRD40	G	0.12	9.45E-06	3.12	(1.88–5.15)	2/34/117	G	T	Drug pathway
rs11140500	9	SLC28A3	Intron	RM11/NTRK2	T	0.01	0.00018	9.94	(2.99–33.09)	0/3/151	T	C	Drug pathway
rs10805074	4	DCK	Intron	MOBK1A/ LOC100128311	A	0.03	0.00039	4.59	(1.98–10.67)	0/9/145	A	G	Drug pathway
rs7684954	4	DCK	Intron	LOC100128311/ LOC727995	A	0.03	0.00039	4.59	(1.98–10.67)	0/9/145	A	G	Drug pathway
rs4593998	11	RRM1	Intron	RRM1/LOC643244	A	0.14	0.00093	2.28	(1.40–3.71)	4/34/116	A	G	Drug pathway
rs6842838	4	DCK	3'UTR of MOB1B	MOB1B/DCK	G	0.05	0.0010	3.21	(1.60–6.45)	0/15/139	G	T	Drug pathway
rs1385985	4	DCK	3'UTR of MOB1B	GRSF1/DCK	C	0.05	0.0010	3.21	(1.60–6.45)	0/15/139	C	T	Drug pathway
rs3754446	1	GSTM1	Near-gene-5 [GSTM5]	GSTM1/GSTM5	G	0.38	0.0012	1.81	(1.26–2.59)	26/66/62	G	T	Drug pathway
rs7689093	4	DCK	Near-gene-5 [DCK]	MOBK1A/DCK	G	0.01	0.0016	5.34	1.89–15.13	0/4/150	G	A	Drug pathway
rs1989983	17	ABCC3	Near-gene-5 [ABCC3]	CACNA1G/ABCC3	A	0.11	0.0017	2.33	(1.37–3.96)	2/30/122	A	G	Drug pathway
rs2301835	17	ABCC3	Synonymous variant in coding of CACNA1G	CACNA1G/ABCC3	T	0.06	0.0029	2.60	(1.39–4.89)	1/16/137	T	C	Drug pathway
rs12515548	5	MSH3	Intron	LOC100128458/ RASGRF2	T	0.13	0.0029	2.07	(1.28–3.35)	6/27/121	T	C	DNA mismatch repair genes
rs2277624	17	ABCC3	Synonymous variant	CACNA1G/ANKRD40	A	0.25	0.0040	1.75	(1.20–2.55)	12/54/87	A	G	Drug pathway
rs11090305	22	GSTT1	Near-gene-5 [CABIN1]	GSTT2/CABIN1	C	0.18	0.0044	1.98	(1.24–3.17)	3/50/101	C	T	Drug pathway
rs7130539	11	RRM1	Intron	STIM1/OR55B1P	C	0.06	0.0047	2.56	(1.33–4.91)	1/16/137	C	T	Drug pathway
rs11031136	11	RRM1	Intergenic	OR55B1P/LOC643244	G	0.06	0.0047	2.56	(1.33–4.91)	1/16/137	G	T	Drug pathway
rs528211	11	SLC22A12	Intergenic	SLC22A11/SLC22A12	G	0.29	0.0048	0.51	(0.31–0.81)	13/62/79	A	G	Other AML study
rs2360872	11	SLC22A12	Intergenic	SLC22A11/SLC22A12	C	0.29	0.0048	0.51	(0.31–0.81)	13/62/79	T	C	Other AML study
rs505802	11	SLC22A12	Near-gene-5 [SLC22A12]	SLC22A11/SLC22A12	A	0.29	0.0048	0.51	(0.31–0.81)	13/62/79	G	A	Other AML study
rs524023	11	SLC22A12	Near-gene-5 [SLC22A12]	SLC22A11/SLC22A12	G	0.29	0.0048	0.51	(0.31–0.81)	13/62/79	A	G	Other AML study
rs9734313	11	SLC22A12	5'UTR	SLC22A11/SLC22A12	T	0.29	0.0048	0.51	(0.31–0.81)	13/62/79	C	T	Other AML study
rs11231825	11	SLC22A12	Synonymous variant	SLC22A11/SLC22A12	C	0.29	0.0048	0.51	(0.31–0.81)	13/62/79	T	C	Other AML study
rs2268166	11	RRM1	Intron	STIM1/OR55B1P	G	0.06	0.0049	2.55	(0.20–0.75)	1/16/136	G	T	Drug pathway
rs11606370	11	SLC22A12	Intron of NRXN2	SLC22A12/RASGRP2	A	0.30	0.0050	0.51	(0.32–0.82)	14/64/76	C	A	Other AML study
rs11191547	10	NT5C2	Intergenic	CNNM2/NT5C2	T	0.31	0.0051	1.71	(1.17–2.49)	15/66/73	T	C	Drug pathway
rs11191549	10	NT5C2	Near-gene-3 [NT5C2]	CNNM2/NT5C2	T	0.31	0.0051	1.71	(1.17–2.49)	15/66/73	T	C	Drug pathway
rs11191553	10	NT5C2	Intron	CNNM2/ LOC100128863	T	0.31	0.0051	1.71	(1.17–2.49)	15/66/73	T	G	Drug pathway
rs10883836	10	NT5C2	Intron	LOC100128863/ LOC729081	C	0.31	0.0051	1.71	(1.17–2.49)	15/66/73	C	T	Drug pathway
rs7095304	10	NT5C2	Intergenic	NT5C2/LOC401648	A	0.31	0.0051	1.71	(1.17–2.49)	15/66/73	A	G	Drug pathway
rs6151816	5	MSH3	Intron	LOC100128458/ RASGRF2	T	0.12	0.0055	2.05	(1.24–3.41)	5/27/122	T	C	DNA mismatch repair genes
rs893006	11	SLC22A12	Intron	SLC22A11/NRXN2	T	0.28	0.0055	0.51	(0.32–0.82)	13/61/80	G	T	Other AML study
rs7818607	8	SLC25A37	Intergenic	SLC25A37/LOC646721	A	0.30	0.0057	1.75	(1.18–2.59)	16/59/79	A	C	Associated with drug cytotoxicity in LBL
rs2853229	8	RRM2B	Intron	NCALD/UBR5	A	0.49	0.0062	1.68	(1.16–2.44)	38/74/42	A	C	Drug pathway
rs8534	8	SLC25A37	Intergenic	SLC25A37/LOC646721	T	0.38	0.0067	1.68	(1.15–2.43)	24/69/61	T	C	Associated with drug cytotoxicity in LBL
rs8079740	17	ABCC3	Intron of CACNA1G	SPATA20/ANCC3	G	0.31	0.0078	0.56	(0.37–0.86)	20/56/78	A	G	Drug pathway
rs757420	17	ABCC3	Intergenic	CACNA1G/ABCC3	T	0.30	0.0079	0.55	(0.36–0.86)	18/56/80	T	C	Drug pathway
rs2010851	12	WNT5B	Near-gene-3 [WNT5B]	WNT5B/ LOC100132548	A	0.30	0.0082	0.52	(0.32–0.84)	13/67/74	C	A	Associated with drug cytotoxicity in LBL
rs4995289	17	P2RX1	Intergenic	P2RX1/ATP2A3	T	0.28	0.0088	0.51	(0.31–0.85)	14/59/81	C	T	Associated with drug cytotoxicity in LBL
rs1516801	17	P2RX1	Intergenic	P2RX1/ATP2A3	G	0.28	0.0088	0.51	(0.31–0.85)	14/59/81	T	G	Associated with drug cytotoxicity in LBL
rs2607662	8	RRM2B	Intron of UBR5	NCALD/UBR5	T	0.46	0.0095	1.64	(1.13–2.37)	31/79/44	T	C	Drug pathway

Abbreviations: AML, acute myeloid leukemia; CI, confidence interval; DFS, disease-free survival; HR, hazards ratio; LBL, lymphoblastoid cell lines; MAF, minor allele frequencies; SNPs, single nucleotide polymorphisms.

The classifications near-gene-5 and near-gene-3 label SNPs that are outside transcribed regions, but within 2000 bp of a transcription region. Near-gene-5 includes upstream promoter region and untranslated 5' mRNA.

the ENCODE Chip-Seq and DNase I peaks (Supplementary Table 2), suggesting that they could have regulatory functions. Interestingly, the GTEx (Genotype-Tissue Expression) eQTL browser (<http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi>) and eQTL browser (<http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>) showed that the SNP in *GSTM1*, rs3754446, is associated with *GSTM1* and *GSTM5* expression levels in the liver,²¹ brain and LCL^{22–24} (Supplementary Table 2).

DISCUSSION

Previous pharmacogenomic studies of AML response to chemotherapy have been limited to a small number of candidate genes.^{2,25–27} Although studies related to genes in drug pathways have been performed, many have not been replicated. In addition, some AML pharmacogenomics studies have been conducted using DNA from

blast cells, which may have included somatic cell mutations in addition to germline polymorphisms.^{28–30}

To determine whether germline genetic variations are associated with AML response, we designed our own pharmacogenomic study in AML patients treated with a chemotherapeutic regimen consisting of cytarabine, etoposide and busulfan, followed by ASCT. Our study was focused primarily on 154 AML patients of European ancestry. The analysis was centered on 42 genes related to the pharmacokinetic and pharmacodynamic pathways of the chemotherapy. A few other genes that had previously been associated with drug response in AML were also included (Supplementary Table 1, Figure 1 and Figure 2). The goals of this study were to identify new associations with DFS in AML patients and to determine whether SNPs previously reported to be associated with AML response could be replicated.

Overall, the most significant SNP in our analysis was the intronic variant in *ABCC3*, rs4148405 (HR = 3.1, $P = 9.5 \times 10^{-6}$). *ABCC3* is a multidrug resistance-associated protein, which is known to transport the etoposide metabolite, etoposide glucuronide.³¹ Several lines of evidence support a role for *ABCC3* in DFS in AML patients. First, in a previous study a promoter variant in *ABCC3*, rs4793665 was associated with a shorter survival time in Israeli AML patients.²⁹ Although this variant was not associated with AML response in the current study, the data support a role of the transporter in response to chemotherapy in AML. Second, following etoposide administration to *Abcc2* $-/-$, *Abcc3* $-/-$ mice, higher etoposide glucuronide levels were observed in the liver,³² consistent with a potentially important role of *ABCC3* in etoposide pharmacokinetics. Third, cell lines transfected with *ABCC3* show greater resistance to etoposide.³³ Finally, higher *ABCC3* expression levels in leukemia cells are associated with poor outcome in children with leukemia.^{34,35}

Data from Regulome database suggest that the *ABCC3* variant, rs4148405, is in a functional location in the genome, as this region demonstrates direct evidence of binding through ChIP-seq studies. Transcription factors that have a role in hepatic gene regulation

(for example, *CEBPB*, *USF1*, *FOXA1*) have DNA response elements within this gene region (see <http://regulome.stanford.edu/snp/chr17/48713567>).^{20,36} On the basis of the results of our primary analysis with DFS, we speculate that the minor allele of rs4148405 is associated with higher expression levels of *ABCC3* in the liver and/or leukemia cells, and thus reduced levels of etoposide in the tumor cells. Although the SNP rs4148405 is found at a MAF of >10% in non-European populations, this SNP was not significantly associated with DFS in the AML patients of non-European ancestry. It is possible that different linkage disequilibrium patterns between rs4148405 and potential causative SNPs may have confounded the analysis.

In addition to a SNP in *ABCC3*, we also identified 23 SNPs in seven other genes (*SLC28A3*, *DCK*, *RRM1*, *GSTM1*, *GSTT1*, *MSH3* and *SLC22A12*) that were associated with DFS in the European AML patients (with $P < 0.005$, Table 2). Interestingly, expression levels of these genes or other SNPs in these genes have been previously associated with response to chemotherapy in AML or other cancers.^{19,29,37–39} Expression levels or SNPs in these genes have also been associated with IC₅₀ values of various chemotherapy agents in cell lines.^{40,41} Genetic polymorphisms in glutathione-S-transferases,

Figure 3 (a–d) Kaplan–Meier estimate of disease-free survival (DFS) stratified by the top SNPs (a) rs4148405 *ABCC3* (b) rs10805074 *DCK* (c) rs3754446 *GSTM1-GSTM5* locus and (d) rs505802 *SLC22A12* genotypes in patients of European ancestry.

Figure 4 Plot showing the association of the genotyped and imputed SNPs with MAF $\geq 1\%$ in the selected top eight candidate genes (*ABCC3*, *DCK*, *SLC28A3*, *SLC22A12*, *MSH3*, *RRM1*, *GSTM1*, *GSTT1*). Each colored diamond (not gray or black) represents a genotyped SNP and the gray/black dots represent imputed SNPs.

such as *GSTT1* and *GSTM1*, have been widely studied for their associations with drug toxicity, drug response and disease risk in leukemia patients.^{29,39,42} Notably, *GSTM1* and *GSTT1* deletions have been implicated in various phenotypes associated with leukemia, including drug response,⁴³ busulfan pharmacokinetics²⁸ and disease risk.⁴⁴ Although we did not examine the effect of the *GSTM1* deletion in this study, we identified several SNPs (genotyped and imputed) in the *GSTM1*-*GSTM5* locus (rs3754446, rs929166 and rs11101989) associated with DFS in individuals of European ancestry that have not been previously reported (Table 2, Figure 2b). One of the SNPs in this locus, rs3754446, was also associated with AML response in the individuals of non-European ancestry (HR = 1.7, $P = 0.028$). Overall, in the entire cohort this minor allele SNP, rs3754446, was significantly associated with shorter DFS (HR = 1.7, $P = 0.00027$). Perusal of eQTL databases (Regulome database, GTEx browser and eQTL browser) suggests that the SNPs in the *GSTM1*-*GSTM5* locus are associated with expression levels of *GSTM1* and/or *GSTM5* in liver,²¹ brain and LCL.^{22–24} Thus, the SNPs in these genes, which are involved in drug metabolism, could affect AML response by affecting the pharmacokinetics of the drugs used in the treatment of AML. We examined 21 AML patients from our overall cohort, where we have their first-dose busulfan area under the curve. Interestingly, we observed a significant association between rs3754446 and reduced busulfan area under the curve in these 21 AML patients ($P = 0.03$, Supplementary Figure 1). The minor allele, G, in rs3754446, was associated with lower busulfan plasma levels (area under the curve), which was consistent with our observation that patients with the G allele had shorter DFS (Figure 3c). As higher busulfan plasma levels have been associated with busulfan liver toxicity,^{14,45} future studies are needed to determine whether the SNPs in *GSTM1*-*GSTM5* are associated with liver toxicity. Collectively, these data suggest that *GSTM1* could have

an important role in determining busulfan drug levels, drug response and/or drug toxicity.

Recently, using a drug-metabolizing enzyme/transporter gene SNP array, a synonymous variant in *SLC22A12* (rs11231825) was found to be associated with response in 94 AML patients treated with a combination drug regimen of gemtuzumab-ozogamicin with fludarabine-cytarabine-idarubicin.¹⁹ Among the SNPs in the drug-metabolizing enzyme/transporter genes that were found to be significantly associated with AML response, we were able to replicate the SNP (rs11231825) in the uric acid transporter, *SLC22A12*. Notably, this synonymous variant, which is in linkage disequilibrium with a SNP upstream of *SLC22A12*, rs505802, has been found in various genome-wide association studies to be associated with uric acid levels.^{46,47} In these genome-wide association studies,^{46,47} the minor allele T has been associated with higher uric acid levels. Although speculative, our study and the previous study,¹⁹ which demonstrated that patients with the T allele have a better response to chemotherapy, suggest that higher uric acid levels may be associated with longer DFS time. In our study, the T allele, which is associated with higher uric acid levels,^{48,49} was associated with longer DFS time (see Figure 3d). Uric acid is a potent antioxidant, and it is possible that higher levels are beneficial for survival in AML patients.

Previous studies have shown that the nucleoside transporter, *SLC28A3* (CNT3) has a role in cytarabine cytotoxicity and resistance.^{40,50,51} In this study, we observed several low-allele frequency variants (MAF 1%) in *SLC28A3* are associated with DFS in AML patients on cytarabine and other chemotherapy. Therefore, we hypothesized that CNT3 may transport cytarabine. Supplementary Figure 2a shows that radiolabeled cytarabine was taken up into CNT3 stably expressing cells, and the uptake (over empty vector cells) was significantly enhanced in cell lines exposed to the equilibrative nucleoside transporter inhibitor S-(4-nitrobenzyl)-6-thioinosine, which reduced background uptake of cytarabine in the cells. Notably, cytarabine uptake decreased significantly in CNT3 stable cells treated with the *SLC28A3* inhibitor (phloridzin) or with both inhibitors (Supplementary Figure 2a). Fludarabine, a known substrate of CNT3 was used as a positive control in this study (Supplementary Figure 2b).⁵² As low-allele frequency variant in CNT3 was found to be significantly associated with DFS (Table 2), we interpret the results with caution. Although our finding that *SLC28A3* transports cytarabine supports the association, functional studies of variants in this region and/or a larger sample size are required to determine whether these uncommon variants are associated with cytarabine response.

DCK has an important role in activating cytarabine to its active metabolite, cytarabine triphosphate. Two tag-SNPs, rs4308342 and rs3775289, in *DCK* that were associated with DFS in our AML patients of European ancestry have been previously associated with the IC₅₀ of another nucleoside analog, gemcitabine, in LCL.⁴⁰ Furthermore, a previous study demonstrated that the level of cytarabine triphosphate in AML blast cells correlates with the ratio in expression levels of the cytarabine-activating enzyme, *DCK* and the -inactivating enzyme, 5'-nucleotidase, cytosolic II (*NT5C2*).⁴⁸ In our study, in addition to the two tag-SNPs in *DCK* associated with DFS in AML patients, five SNPs in the *NT5C2* region were associated with DFS (Table 2). Although several of these SNPs are eQTLs (see Supplementary Table 2), further studies are required to determine whether these SNPs have important roles in determining the levels of cytarabine triphosphate in AML blast cells. Other genes, which have important roles in the cytarabine pharmacodynamics pathway are ribonucleotide reductase, *RRM1* and *RRM2*, which are considered

targets of nucleoside drugs such as cytarabine. The role of this enzyme is to regulate intracellular pools of ribonucleotides, such as deoxycytidine triphosphate, which is important in building blocks for DNA replication. Studies have shown that AML blasts cells with high levels of deoxycytidine triphosphate are resistant to cytarabine, and that there is a significant correlation between *RRM1* and *RRM2* gene expression levels and deoxycytidine triphosphate levels after cytarabine treatment in AML blast cells.¹⁵ Consistent with previous studies showing that SNPs in *RRM1* are associated with response or toxicity to gemcitabine-based chemotherapy in lung and breast cancer patients,^{49,53} our findings suggest that the SNPs in *RRM1* are associated with AML response to chemotherapy that include cytarabine.

Although our current association analysis supports the important roles of drug pathway genes, mainly transporters and enzymes, in AML response, we also selected eight genes that have been associated with cytarabine or etoposide IC₅₀ values in LCL.^{16,17} A few SNPs in the eight genes were significantly associated with response in AML patients, suggesting that genes identified in *in vitro* assays in LCLs may also be important predictors of *in vivo* drug response in AML patients.^{16,17}

In summary, in this genetic association study of DFS in AML patients, we identified polymorphisms that have not been previously associated with AML response, including SNPs in *ABCC3*, *DCK*, *GSTM1*, *MSH3*, *NT5C2*, *RRM1* and *SLC28A3*. A SNP in *ABCC3*, rs4148405, which remained significant after multiple testing, was associated with DFS, suggesting an important role of *ABCC3* in determining etoposide levels in the liver and other tissues, and hence AML response. Many of the significant SNPs or their tag-SNPs were eQTLs or located in functional regions in the genome. Finally, we determined for the first time that *SLC28A3* (CNT3) transported cytarabine into cells, suggesting an important role of this transporter in cytarabine cytotoxicity.

ACKNOWLEDGEMENTS

This research was supported in part by a grant from NIH NIGMS (GM61390). This study was supported by the NIH Pharmacogenomics Research Network (PGRN)-RIKEN Center for Genomic Medicine (CGM) Strategic Alliance. We are grateful to the UCSF AML Tissue Bank staff (Ms Joy Cruz, Ms Christine Cheng). SWY thanks Chris Gignoux and Pär Mattson for their help in creating plots in the R-project.

- Daver, N. & Cortes, J. Molecular targeted therapy in acute myeloid leukemia. *Hematology* **17** (Suppl 1), S59–S62 (2012).
- Patel, J. P., Gonen, M., Figueroa, M. E., Fernandez, H., Sun, Z., Racevskis, J. *et al.* Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N. Engl. J. Med.* **366**, 1079–1089 (2012).
- Buccisano, F., Maurillo, L., Del Principe, M. I., Del Poeta, G., Sconocchia, G., Lo-Coco, F. *et al.* Prognostic and therapeutic implications of minimal residual disease detection in acute myeloid leukemia. *Blood* **119**, 332–341 (2012).
- Martelli, M. P., Sportoletti, P., Tiacci, E., Martelli, M. F. & Falini, B. Mutational landscape of AML with normal cytogenetics: Biological and clinical implications. *Blood Rev.* **27**, 13–22 (2013).
- Kroger, N., Brand, R., van Biezen, A., Cahn, J. Y., Slavina, S., Blaise, D. *et al.* Autologous stem cell transplantation for therapy-related acute myeloid leukemia and myelodysplastic syndrome. *Bone Marrow Transplant.* **37**, 183–189 (2006).
- Novitzky, N., Thomas, V., du Toit, C. & McDonald, A. Is there a role for autologous stem cell transplantation for patients with acute myelogenous leukemia? A retrospective analysis. *Biol. Blood Marrow Transplant.* **17**, 875–884 (2011).
- Linker, C. A., Damon, L. E., Ries, C. A., Navarro, W. A., Case, D. & Wolf, J. L. Autologous stem cell transplantation for advanced acute myeloid leukemia. *Bone Marrow Transplant.* **29**, 297–301 (2002).
- Linker, C. A., Owzar, K., Powell, B., Hurd, D., Damon, L. E., Archer, L. E. *et al.* Auto-SCT for AML in second remission: CALGB study 9620. *Bone Marrow Transplant.* **44**, 353–359 (2009).
- Linker, C. A., Ries, C. A., Damon, L. E., Sayre, P., Navarro, W., Rugo, H. S. *et al.* Autologous stem cell transplantation for acute myeloid leukemia in first remission. *Biol. Blood Marrow Transplant.* **6**, 50–57 (2000).
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
- Yang, J., Bogni, A., Schuetz, E. G., Ratain, M., Dolan, M. E., McLeod, H. *et al.* Etoposide pathway. *Pharmacogenet. Genomics* **19**, 552–553 (2009).
- Li, L., Schaid, D. J., Fridley, B. L., Kalari, K. R., Jenkins, G. D., Abo, R. P. *et al.* Gemcitabine metabolic pathway genetic polymorphisms and response in patients with non-small cell lung cancer. *Pharmacogenet. Genomics* **22**, 105–116 (2012).
- Hassan, M. & Andersson, B. S. Role of pharmacogenetics in busulfan/cyclophosphamide conditioning therapy prior to hematopoietic stem cell transplantation. *Pharmacogenomics* **14**, 75–87 (2013).
- Abbasi, N., Vadnais, B., Knutson, J. A., Blough, D. K., Kelly, E. J., O'Donnell, P. V. *et al.* Pharmacogenetics of intravenous and oral busulfan in hematopoietic cell transplant recipients. *J. Clin. Pharmacol.* **51**, 1429–1438 (2011).
- Lamba, J. K. Genetic factors influencing cytarabine therapy. *Pharmacogenomics* **10**, 1657–1674 (2009).
- Huang, R. S., Duan, S., Bleibel, W. K., Kistner, E. O., Zhang, W., Clark, T. A. *et al.* A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. *Proc. Natl Acad. Sci. USA* **104**, 9758–9763 (2007).
- Hartford, C. M., Duan, S., Delaney, S. M., Mi, S., Kistner, E. O., Lamba, J. K. *et al.* Population-specific genetic variants important in susceptibility to cytarabine arabinoside cytotoxicity. *Blood* **113**, 2145–2153 (2009).
- Hewish, M., Martin, S. A., Elliott, R., Cunningham, D., Lord, C. J. & Ashworth, A. Cytosine-based nucleoside analogs are selectively lethal to DNA mismatch repair-deficient tumour cells by enhancing levels of intracellular oxidative stress. *Br. J. Cancer* **108**, 983–992 (2013).
- Iacobucci, I., Lonetti, A., Candoni, A., Sazzini, M., Papayannidis, C., Formica, S. *et al.* Profiling of drug-metabolizing enzymes/transporters in CD33+ acute myeloid leukemia patients treated with Gemtuzumab-Ozogamicin and Fludarabine, Cytarabine and Idarubicin. *Pharmacogenomics J.* (e-pub ahead of print 15 May 2012; doi:10.1038/tpj.2012.13).
- Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
- Schadt, E. E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P. Y. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
- Stranger, B. E., Nica, A. C., Forrest, M. S., Dimas, A., Bird, C. P., Beazley, C. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
- Veyrieras, J. B., Kudravalli, S., Kim, S. Y., Dermizakis, E. T., Gilad, Y., Stephens, M. *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* **4**, e1000214 (2008).
- Gibbs, J. R., van der Brug, M. P., Hernandez, D. G., Traynor, B. J., Nalls, M. A., Lai, S. L. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6**, e1000952 (2010).
- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
- Gaidzik, V. I., Paschka, P., Spath, D., Haddank, M., Kohne, C. H., Germing, U. *et al.* TET2 mutations in acute myeloid leukemia (AML): results from a comprehensive genetic and clinical analysis of the AML study group. *J. Clin. Oncol.* **30**, 1350–1357 (2012).
- Paschka, P., Marcucci, G., Ruppert, A. S., Whitman, S. P., Mrozek, K., Maharry, K. *et al.* Wilms' tumor 1 gene mutations independently predict poor outcome in adults with cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. *J. Clin. Oncol.* **26**, 4595–4602 (2008).
- Kim, S. D., Lee, J. H., Hur, E. H., Lee, J. H., Kim, D. Y., Lim, S. N. *et al.* Influence of GST gene polymorphisms on the clearance of intravenous busulfan in adult patients undergoing hematopoietic cell transplantation. *Biol. Blood Marrow Transplant.* **17**, 1222–1230 (2011).
- Muller, P., Asher, N., Heled, M., Cohen, S. B., Risch, A. & Rund, D. Polymorphisms in transporter and phase II metabolism genes as potential modifiers of the predisposition to and treatment outcome of de novo acute myeloid leukemia in Israeli ethnic groups. *Leuk. Res.* **32**, 919–929 (2008).
- Emadi, A. & Karp, J. E. The clinically relevant pharmacogenomic changes in acute myelogenous leukemia. *Pharmacogenomics* **13**, 1257–1269 (2012).
- Zelcer, N., Saeki, T., Reid, G., Beijnen, J. H. & Borst, P. Characterization of drug transport by the human multidrug resistance protein 3 (ABCC3). *J. Biol. Chem.* **276**, 46400–46407 (2001).
- Lagas, J. S., Fan, L., Wagenaar, E., Vlaming, M. L., van Tellingen, O., Beijnen, J. H. *et al.* P-glycoprotein (P-gp/Abcb1), Abcc2, and Abcc3 determine the pharmacokinetics of etoposide. *Clin. Cancer Res.* **16**, 130–140 (2010).
- Belinsky, M. G., Dawson, P. A., Shchavleva, I., Bain, L. J., Wang, R., Ling, V. *et al.* Analysis of the *in vivo* functions of Mrp3. *Mol. Pharmacol.* **68**, 160–168 (2005).
- Steinbach, D., Wittig, S., Cario, G., Viehmann, S., Mueller, A., Gruhn, B. *et al.* The multidrug resistance-associated protein 3 (MRP3) is associated with a poor outcome in childhood ALL and may account for the worse prognosis in male patients and T-cell immunophenotype. *Blood* **102**, 4493–4498 (2003).
- Steinbach, D., Lengemann, J., Voigt, A., Hermann, J., Zintl, F. & Sauerbrey, A. Response to chemotherapy and expression of the genes encoding the multidrug

- resistance-associated proteins MRP2, MRP3, MRP4, MRP5, and SMRP in childhood acute myeloid leukemia. *Clin. Cancer Res.* **9**, 1083–1086 (2003).
- 36 Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- 37 Kim, K. I., Huh, I. S., Kim, I. W., Park, T., Ahn, K. S., Yoon, S. S. *et al.* Combined interaction of multi-locus genetic polymorphisms in cytarabine arabinoside metabolic pathway on clinical outcomes in adult acute myeloid leukaemia (AML) patients. *Eur. J. Cancer* **49**, 403–410 (2013).
- 38 Shi, J. Y., Shi, Z. Z., Zhang, S. J., Zhu, Y. M., Gu, B. W., Li, G. *et al.* Association between single nucleotide polymorphisms in deoxycytidine kinase and treatment response among acute myeloid leukaemia patients. *Pharmacogenetics* **14**, 759–768 (2004).
- 39 Xiao, Z., Yang, L., Xu, Z., Zhang, Y., Liu, L., Nie, L. *et al.* Glutathione S-transferases (GSTT1 and GSTM1) genes polymorphisms and the treatment response and prognosis in Chinese patients with de novo acute myeloid leukemia. *Leuk. Res.* **32**, 1288–1291 (2008).
- 40 Li, L., Fridley, B. L., Kalari, K., Jenkins, G., Batzler, A., Weinshilboum, R. M. *et al.* Gemcitabine and arabinosylcytosin pharmacogenomics: genome-wide association and drug response biomarkers. *PLoS one* **4**, e7765 (2009).
- 41 Mitra, A. K., Crews, K. R., Pounds, S., Cao, X., Feldberg, T., Ghodke, Y. *et al.* Genetic variants in cytosolic 5'-nucleotidase II are associated with its expression and cytarabine sensitivity in HapMap cell lines and in patients with acute myeloid leukemia. *J. Pharmacol. Exp. Ther.* **339**, 9–23 (2011).
- 42 Mossallam, G. I., Abdel Hamid, T. M. & Samra, M. A. Glutathione S-transferase GSTM1 and GSTT1 polymorphisms in adult acute myeloid leukemia; its impact on toxicity and response to chemotherapy. *J. Egypt. Natl. Canc. Inst.* **18**, 264–273 (2006).
- 43 Voso, M. T., Hohaus, S., Guidi, F., Fabiani, E., D'Alo, F., Groner, S. *et al.* Prognostic role of glutathione S-transferase polymorphisms in acute myeloid leukemia. *Leukemia* **22**, 1685–1691 (2008).
- 44 Das, P., Shaik, A. P. & Bammidi, V. K. Meta-analysis study of glutathione-S-transferases (GSTM1, GSTP1, and GSTT1) gene polymorphisms and risk of acute myeloid leukemia. *Leuk. Lymphoma* **50**, 1345–1351 (2009).
- 45 Zhang, H., Graiser, M., Hutcherson, D. A., Dada, M. O., McMillan, S., Ali, Z. *et al.* Pharmacokinetic-directed high-dose busulfan combined with cyclophosphamide and etoposide results in predictable drug levels and durable long-term survival in lymphoma patients undergoing autologous stem cell transplantation. *Biol. Blood. Marrow Transplant.* **18**, 1287–1294 (2012).
- 46 Takeuchi, F., Yamamoto, K., Isono, M., Katsuya, T., Akiyama, K., Ohnaka, K. *et al.* Genetic Impact on Uric Acid Concentration and Hyperuricemia in the Japanese Population. *J. Atheroscler. Thromb.* **20**, 351–367 (2012).
- 47 Kolz, M., Johnson, T., Sanna, S., Teumer, A., Vitart, V., Perola, M. *et al.* Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet.* **5**, e1000504 (2009).
- 48 Yamauchi, T., Negoro, E., Kishi, S., Takagi, K., Yoshida, A., Urasaki, Y. *et al.* Intracellular cytarabine triphosphate production correlates to deoxycytidine kinase/cytosolic 5'-nucleotidase II expression ratio in primary acute myeloid leukemia cells. *Biochem. Pharmacol.* **77**, 1780–1786 (2009).
- 49 Kim, S. O., Jeong, J. Y., Kim, M. R., Cho, H. J., Ju, J. Y., Kwon, Y. S. *et al.* Efficacy of gemcitabine in patients with non-small cell lung cancer according to promoter polymorphisms of the ribonucleotide reductase M1 gene. *Clin. Cancer Res.* **14**, 3083–3088 (2008).
- 50 Errasti-Murugarren, E., Pastor-Anglada, M. & Casado, F. J. Role of CNT3 in the transepithelial flux of nucleosides and nucleoside-derived drugs. *J. Physiol.* **582**, 1249–1260 (2007).
- 51 Sarkar, M., Han, T., Damaraju, V., Carpenter, P., Cass, C. E. & Agarwal, R. P. Cytosine arabinoside affects multiple cellular factors and induces drug resistance in human lymphoid cells. *Biochem. Pharmacol.* **70**, 426–432 (2005).
- 52 Badagnani, I., Chan, W., Castro, R. A., Brett, C. M., Huang, C. C., Stryke, D. *et al.* Functional analysis of genetic variants in the human concentrative nucleoside transporter 3 (CNT3; SLC28A3). *Pharmacogenomics J.* **5**, 157–165 (2005).
- 53 Rha, S. Y., Jeung, H. C., Choi, Y. H., Yang, W. I., Yoo, J. H., Kim, B. S. *et al.* An association between RRM1 haplotype and gemcitabine-induced neutropenia in breast cancer patients. *Oncologist* **12**, 622–630 (2007).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)