JPET Fast Forward. Published on September 12, 2007 as DOI: 10.1124/jpet.107.128595

JPET #128595

Pharmacogenetics of Deoxycytidine kinase: identification and characterization of novel genetic variants.

Jatinder K. Lamba, Kristine Crews, Stanley Pounds, Erin G. Schuetz, Jessica Gresham, Varsha Gandhi, William Plunkett, Jeffrey Rubnitz, and Raul Ribeiro.

Department of Pharmaceutical Sciences (J.K.L., K.C., E.G.S., J.G.) , Department of Biostatistics (S.P.), Department of Oncology (J.R, R.R.), St. Jude Children's Research Hospital, Memphis, TN. Department of Experimental Therapeutics (V.G., W.P.), MD Anderson Cancer Center, Houston, TX.

Running title: Pharmacogenetics of DCK.

Corresponding author:
Jatinder Lamba, PhD
St. Jude Children's Research Hospital
332 North Lauderdale St.
Memphis, TN, 38105.
Email: jatinder.lamba@stjude.org
Phone: (901) 495 5480
Fax: (901) 495 3125

Text pages: 24
Tables: 3
Figures: 6
References: 20
Abstract: 250 words
Introduction: 571 words
Discussion: 1195 words

Nonstandard abbreviations:
5'UTR: 5' untranslated region; Ara-C: 1-β-D-arabinofuranosyl-cytosine (Cytarabine); CdA:Cladribine; CEPH: Centre d' Etude du Polymorphisme Humain; DCK: Deoxycytidine kinase; ESE: Exonic splicing enhancer; IVS: Intervening sequence (intron); HapMap: Haplotype Map (of human genome); mt: mutant; LD: Linage disequilibrium SNPs: Single nucleotide Polymorphisms; WT: wildtype; YRI: Yoruba people in Ibadan, Nigeria

Recommended section assignment: Metabolism, Transport, and Pharmacogenomics

Abstract

Deoxycytidine kinase (DCK) is a rate limiting enzyme in the activation of nucleoside analogs such as ara-C, gemcitabine, clofarabine etc. The present study was undertaken to identify and to determine the functional consequences of genetic variants in DCK. We sequenced 1.5 kb of the DCK proximal promoter and all 7 coding exons in HapMap panels (n=90 each) with European (CEPH) or African (YRI) ancestry. Sixty-four genetic polymorphisms, including 3 non-synonymous coding changes (Ile24Val, Ala119Gly, and Pro122Ser) were identified. Compared to DCK-wildtype (WT) protein, the activity of the recombinant DCK24Val, DCK119Gly, and DCK122Ser proteins was 85±5%, 66±3%, and 43±4%, respectively. DCK119Gly and DCK122Ser mutants had lower Km ($p<0.01$) and Vmax ($p<0.001$) as compared to DCK-WT protein. Lymphoblast cell lines from subjects heterozygous for the coding changes had significantly lower DCK activity as compared to homozygous WT subjects. Ethnic differences were observed, with African ancestry subjects demonstrating significantly higher DCK mRNA expression as compared to subjects with European ancestry. In both CEPH and YRI subjects, the C allele of a 3'UTR SNP (*35708 C>T)*, was significantly associated with lower DCK mRNA expression. This SNP was strongly linked with other intronic SNPs forming a major haplotype block in both ethnic groups. In an exploratory analysis, the *35708C* allele was also associated with lower blast ara-CTP levels in AML patients receiving ara-C as continuous infusion. These results suggest that genetic variation in DCK influences its activity and expression and may predict the variability observed in intracellular levels of the ara-C active metabolite, ara-CTP.

Introduction:

Cytarabine (ara-C) has been part of the chemotherapeutic regimens used in the treatment of AML for over 40 years (Wang, et al., 1970). Like other nucleoside analogs, ara-C is a prodrug that requires extensive intracellular phosphorylation for activation to its active metabolite ara-C-5'-triphosphate (ara-CTP). Ara-C phosphorylation to ara-C-5'-monophosphate (ara-CMP) by deoxycytidine kinase (DCK) is the rate-limiting step in its activation. Ara-CMP is then further phosphorylated by mono and di-phophokinases to ara-CTP. Ara-CTP incorporates into DNA in place of deoxycytidine triphosphate (dCTP), resulting in chain termination, thereby blocking DNA synthesis and causing leukemic cell death (Kufe, et al., 1980;Major, et al., 1981;Raza, et al., 1992).

In vitro studies have shown that ara-C sensitive cells accumulate higher intracellular concentrations of ara-CTP than do resistant cells (Kufe, et al., 1984). Development of resistance to ara-C and other nucleoside analogs has been associated with reduced ara-C uptake into the cell due to lower expression of human equilibrative nucleoside transporter (hENT1), reduced ara-C activation due to lower expression of DCK, or increased ara-C inactivation due to higher expression of the inactivating enzymes cytidine deaminase (CDA), and NT5C2 (5'nucleotidase) (reviewed by (Cros, et al., 2005).

Using H9 lymphoid cell lines, Sarkar et al. have shown that compared to the ara-C sensitive cell line, the resistant cell line accumulated less ara-CTP and had significantly lower mRNA and/or protein expression of DCK and hENT1 (Sarkar, et al., 2005). When treated with ara-C, patients whose leukemic cells had higher DCK expression, demonstrated longer event-free survival than patients whose leukemic cells had lower DCK expression (Galmarini, et al., 2002). Further, in xenograft models, the pretreatment levels of DCK are also related to gemcitabine sensitivity in solid tumors of different origins (Kroep, et al., 2002).

Development of resistance to ara-C is thus multi-factorial, with DCK appearing to play a distinct role. The activity and expression of DCK varies widely in normal and malignant cells and tissues. There is approximately a 50-fold variation in DCK expression in patient leukemic cells (Kakihara, et al., 1998), 35-fold variation in DCK mRNA in primary AML cells, 36-fold variation of DCK mRNA in normal liver tissue, and

150-fold variation of DCK mRNA in human liver metastases of colorectal cancer origin (van der Wilt, et al., 2003).

The human DCK gene is located on chromosome 4q13.3-12.1 and has 7 exons. Recently, the DCK gene and promoter have been screened for sequence variations in a Chinese population (Shi, et al., 2004). Two promoter variants, -360C>G, and -201C>T (rs2306744) that are in strong linkage disequilibrium (LD) were associated with transcriptional activation of reporter constructs. No coding variant was identified in this Chinese population (Shi, et al., 2004). Recently, 6 novel genetic variants in the DCK gene (-243G>T, -135G>C, 261G>A, 364C>T(P121S), 727A>C (K242P), and Int6 T>A) has been reported in Caucasians (Joerger, et al., 2006). However the functional consequences of these changes are not known.

The present study was designed with the following aims: 1) to identify new functionally significant genetic variants/SNPs in the DCK gene by re-sequencing the genomic DNA from subjects with European or African ancestries; 2) to determine the association of SNPs from the International HapMap Project (HapMap; www.hapmap.org) and those identified in the present study with DCK mRNA expression; 3) to determine the effect of any coding variants on DCK activity and kinetics, and; 4) to determine the association between DCK genetic variants and ara-CTP concentration in leukemic cells of children with AML undergoing treatment with ara-C.

Materials and Methods:

Reagents

[8-^{3}H]-2-Chloro-2'-deoxyadenosine (^{3}H-Cladribine, CdA) and Cladribine were obtained from Moravek Biochemicals, (Brea, CA) and Sigma Aldrich Inc. (Atlanta, GA), respectively. PEI-Cellulose TLC sheets were obtained from Fisher Scientific (Atlanta, GA). Trizol tri-reagent used for RNA isolation, the SuperScript II First Strand cDNA Synthesis Kit, and the pET Directional Cloning Kit were obtained from Invitrogen (Carlsbad, CA). The Expand Hi Fidelity PCR system was obtained from Roche (Indianapolis, IN). Anti-rabbit DCK-C-ter antibody and DCK blocking peptide were

obtained from Abgent (San Diego, CA). All other reagents used were of a quality suitable for use in molecular biology studies.

Study population for DCK

For the present study we used Epstein-Barr virus-transformed B-lymphoblastoid HapMap cell lines derived from 30 Centre d' Etude du Polymorphisme Humain (CEPH) trios (2 parents and a child) (n = 90, European descent) and 30 Yoruba trios (n = 90, African descent, referred as YRI) to identify genetic variants in DCK. The purpose of using the same cell lines that have been used in the International HapMap project was to allow us to utilize the genotype data generated as part of the HapMap project.

Samples of the lymphoblast cell lines YRI (n = 90) and CEPH (n = 90) were obtained from the non-profit Coriell Institute for Medicine (Camden, NJ; www.ccr.coriell.org). Cell lines were grown in an RPMI-1640 medium supplemented with 2 mM L-Glutamine (BioWhittaker; Walkersfield, MD) and 15% heat-inactivated serum at 37°C under 5% CO_2. The DNA, RNA, and cytoplasmic fractions were extracted from these cell lines using standard protocols. Genomic DNA was used to discover novel genetic variants in the DCK gene, RNA samples were used to quantitate expression of DCK by Taqman real-time PCR, and the cytoplasmic fractions were used to determine DCK activity, as described below.

Identification of sequence variations in the DCK gene

All the coding exons, intron 1, and 1.5 kb of the 5'UTR of the DCK gene were PCR-amplified using primers and conditions listed in Table 1. Amplification was carried out in a 1 x PCR buffer using 10 ng of genomic DNA, 10 pmol each of forward and reverse primers, 0.2 mM dNTPs, and 1.5 units of Taq polymerase (Expand High Fidelity PCR System; Roche, Altanta, GA). Prior to sequencing, unincorporated nucleotides and primers were removed by incubation with Shrimp Alkaline Phosphatase and Exonuclease I (USB; Cleveland, OH) for 30 min at 37°C, followed by inactivation at 80°C for 15 min.

Sequencing was carried out with an ABI Prism 3700 Automated Sequencer using the PCR primers or internal primers (sequence available on request). Sequences were

assembled using the Phred-Phrap-Consed package (University of Washington; Seattle, WA; http://droog.mbt.washington.edu/PolyPhred.html), which automatically detects the presence of heterozygous single nucleotide substitutions by fluorescence-based sequencing of PCR products (Nickerson, et al., 1997)

DCK expression constructs

Full-length DCK cDNA was amplified from one of the samples representing the WT sequence (Gene bank accession # NT_000788) and was cloned into a pET101 expression vector using a Champion pET Directional TOPO Expression Kit from Invitrogen (Catalog # K101-1), as per the manufacturer's instructions. Site-directed mutagenesis was performed using a QuikChange SDM Kit (Stratagene; LaJolla, CA) to create the DCK-24mt, DCK-119mt, and DCK-122mt (positions refer to amino acid) mutant expression constructs. The empty pET101 vector, pET101-DCK WT, pET101-DCK 24mt, pET101-DCK 119mt, and pET101-DCK 122mt expression vectors were then transformed into the BL21 star *E. coli* provided with the kit, as per manufacturer's instructions. Expression of DCK was induced by 1 mM IPTG and cells were harvested after 3 h. The cell pellet was suspended in 50 mM tris-(hydroxymethyl)-amino methane (TRIS)/HCl buffer (pH 7.4) containing 4 mM DTT and the soluble fraction prepared as per manufacturer's instructions. The soluble fraction was checked for DCK expression by western blotting before proceeding to activity assays.

Western Blotting

The amount of the bacterially-expressed DCK recombinant protein levels among different expression constructs was assessed by performing western blotting. Cytoplasmic fractions from bacterially-expressed DCK recombinant proteins were electrophoresed using 12.5 % sodium dodecyl sulphate (SDS) polyacrylamide gels and transferred by electroblotting to Hybond nitrocellulose membranes (Hybond ECL; Amersham Biosciences Ltd). The blots were incubated overnight with blocking buffer containing 5% BSA in phosphate-buffered saline, containing 0.1% Tween 20 and probed with rabbit anti-Human DCK polyclonal C-ter antibody (Abgent) for 1 h at room temperature, followed by incubation with anti-rabbit secondary antibody conjugated to

horse radish peroxidase. Immune-complexes formed were visualized by enhanced chemiluminescence reaction.

The specificity of the antibody was confirmed by pre-incubating the primary antibody with DCK blocking-peptide (Abgent) prior to probing, as per the manufacturer's instructions.

DCK activity assays

DCK activity of the recombinant WT and mutant DCK protein expressed in BL21 cells, as well as in selective CEPH and YRI cell lines, was determined using ^{3}H-Cladribine (CdA) as a substrate. The DCK activity assay was based on the method originally described by Arner and colleagues (Arner, et al., 1992) and as modified by van der Wilt et al (van der Wilt, et al., 2003). Briefly, cell pellets (at least 1.5 x 10^{7} for lymphoblast cell lines) were resuspended in 50 mM tris-(hydroxymethyl)-amino methane (TRIS)/HCl buffer (pH 7.4) containing 4 mM DTT, were disrupted by sonication and centrifuged at 10,000 x g for 15 min. The supernatant was used for enzyme assays. Protein was estimated by the Bio-Rad protein assay using bovine serum albumin as the standard. 25 μl of supernatant from lymphoblast cell lines, or 10 μl of the bacterially expressed recombinant protein was added along with 10 μl of 50 mM MgATP/100mM NaF, ^{3}H-CdA (128 mmol; specific activity 7.8 Ci/mmol)—for kinetic studies different concentrations of 3H-CdA were used—in a total volume of 50 μl. The reaction was carried out at 37°C for 30 min, and then was terminated by heating to 94°C for 4 min. The reaction mix was then centrifuged, and 5 μl of the supernatant was spotted on PEI cellulose TLC sheets. The monophosphate product was separated from the substrate as described earlier (van der Wilt, et al., 2003). Enzyme activity was expressed as pmol or nmol CdA-MP formed/mg protein · h^{-1}.

Real time quantitation of DCK mRNA

Total RNA was isolated from the lymphoblast cell lines using Trizol Reagent (Invitrogen). The quality of RNA was assessed using an Agilent Bioanalyzer before performing real-time quantification. First-strand cDNA was prepared using oligo-dT primers (SuperScript RT-PCR System; Invitrogen). TaqMan real-time PCR quantitation

of DCK and the endogenous control GAPDH were carried out using primers and probes from Applied Biosystem's TaqMan Gene Expression Assays. cDNA was analyzed in duplicate by TaqMan real-time PCR on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems; Foster City, CA). Quantitation was normalized to the endogenous control GAPDH. Standard curves were prepared for both the target gene and GAPDH using serially diluted cDNA for a highly expressed sample (in triplicate). Real-time values were determined using the comparative C_T ($\Delta\Delta C_T$) method. To determine the quantity of the DCK transcripts present the Ct values were first normalized using GAPDH as control (ΔCt = Ct, DCK – Ct, GAPDH). The relative concentration was determined by 2–ΔΔCt method.

Genotyping of DCK SNPs in genomic DNA of AML patients

Genomic DNA from 55 children, newly diagnosed with AML, was analyzed for the coding genetic variants in exon 3 and exon 7, and the 3'UTR SNPs of DCK (includes SNP representing Block 1). These patients were enrolled in the St. Jude AML97 protocol. The eligibility for the enrollment and treatment plan of the protocol has been published elsewhere (Crews, et al., 2002).

Patients were randomly assigned to receive either a daily short infusion or a continuous infusion of ara-C. Patients receiving the short daily infusion were given a daily 2-hour intravenous (IV) infusion of ara-C (500 mg/m^2) on each of 5 consecutive days. Patients in the continuous infusion arm of the study received ara-C (500 mg/m^2 · d^{-1}) as a 120-hour continuous infusion. Bone marrow aspirates were obtained at 2 h or 10 h following the start of ara-C infusion, from patients receiving short or continuous infusion, respectively. Leukemic cells were separated by Ficoll-Hypaque density-gradient centrifugation and ara-CTP levels were determined using High Performance Liquid Chromatography as described earlier (Crews, et al., 2002). An Interim analysis of this study has been reported (Crews, et al., 2002).

The study design and pharmacologic investigations were approved by the St. Jude Children's Research Hospital Institutional Review Board. Written informed consent was obtained from patients, parents, or guardians (as appropriate) before enrollment into the

study. The present retrospective study was approved by the Institution's Ethics Committee.

Statistical Analysis

SNP-specific comparisons of DCK expression in HapMap cell lines across genotype were performed using an analysis of variance (ANOVA) model that used a Toeplitz correlation structure (Wolfinger R.D., 1996) with diagonal bands that allowed a child's measurements to correlate with their parents' measurements but assumed all other pairs of measurements to be independent. A similar model was used to explore the association of DCK expression with race and gender. DCK expression levels were log-transformed in these analyses to better represent them with a normal distribution and to stabilize variance.

We used the Wilcoxon rank-sum test to perform comparisons of DCK activity and DCK-activity-to-expression ratio in HapMap cell lines. Additionally, Spearman's rank-based correlation coefficient was used to measure the association of DCK mRNA and DCK activity in HapMap cell lines. A z-test was used to compare the correlation of homozygous subjects to that of subjects heterozygous in at least 1 allele. The z-statistic was computed by dividing the difference of the observed correlations by the square root of the sum of their estimated standard errors.

Welch's t-test was used to compare DCK activity of the wildtype and mutant recombinant DCK protein. Graphpad prism software was used to calculate apparent Km and Vmax values (and their standard errors) for recombinant wildtype and mutant DCK proteins (Graphpad software; San Diego, CA). The estimates and standard errors were used in a Welch t-test to perform pairwise comparisons of Km and Vmax between wildtype and each mutant genotype. The rank-sum test was used to perform within-arm SNP-specific comparisons of expression across genotypes for AML97 samples. All tests were two-sided; no multiple testing adjustments were performed in these exploratory analyses. All statistical calculations (except for Km and Vmax values) were performed using SAS version 9 (SAS Institute; Cary, NC).

Results:

Novel genetic variants identified in DCK gene

All the 7 coding exons of DCK and 1.5 kb of the proximal promoter were sequenced in 180 samples: 90 with European ancestry (CEPHs) and 90 with African Ancestry (YRI). A total of 64 variants including 2 insertion/deletion polymorphisms were identified in the CEPH and YRI samples. Twelve were present in the promoter, 14 in 3'UTR, 34 were intronic and 4 were in the coding region (Fig. 1A). Three out of 4 coding changes were nonsynonymous resulting in Ile24Val (*A70G*, exon 1), Ala119Gly (*C28680G*, exon 3), and Pro122Ser (*C28688T*, exon 3) changes. Ile24Val and Ala119Gly changes were specific to the group with African ancestry (YRI), whereas Pro122Ser was identified in both CEPH and YRI ethnic groups.

Genotype data on 37 variants is available at the DCK locus from the International HapMap project, of which 3 are in 3'UTR and 34 are intronic. There is an overlap of 9 variants between the present study and the HapMap genotype data. Table 2 lists all the genetic variants identified in the present study along with their frequency in the 2 ethnic groups.

Analysis for Linkage disequilibrium was performed using genotype information on all the SNPs identified in the present study and from the HapMap project (www.HAPMAP.org). HaploView was used to calculate D', LOD and r^2 values for pair wise combinations of all genetic polymorphisms identified, variants with the minimum allele frequency of less than 0.05 were excluded from the analysis. LD plots demonstrating the linkage pattern (r^2 values) at the DCK locus in both European (CEPH) and African (YRI) ethnic groups were generated using Haploview and are shown in Fig. 1B. The strong LD is indicated by black (r^2 =1), intermediate LD with shades of grey ($0<r^2<1$) and no evidence of LD with white ($r^2=0$). Within CEPH samples there was strong LD between multiple SNPs at the DCK locus. Based on the LD plot and results from tagger program (part of Haploview used for tag SNP selection) we were able to group the strongly linked SNPs (r^2 >0.8) in CEPHs into 4 groups. Within YRI samples, there were more haplotypes resulting in 9 distinct groups of SNPs, as shown in Fig. 1B. Most of the SNPs in Group1/Block1 are common to both the CEPH and YRI ethnic groups. Interestingly as is indicated in Fig. 1B, the frequency Block 1

SNPs in CEPH was 0.05 as compared to 0.77 in YRI samples. This difference in the frequency of the SNPs indicates that ethnic differences in allele frequencies could contribute to the phenotypic differences.

Functional characterization of coding genetic variants in DCK

We compared the amino acid sequence around the coding changes observed in DCK (IIe24Val, Ala119Gly, and Pro122Ser) in 5 difference species: cow, human, monkey (macaque), mouse, and rat. These DCK sequences were obtained from Genebank, NCBI. As is shown in Fig. 2A, the DCK amino acids were highly conserved across the 5 species. Ala119 and Pro122 amino acid residues are present in close proximity to an ERS motif (amino acids 127-129) which is present in the active site of DCK. Arg128 in the ERS motif has been shown to interact with ara-C (Johnsamuel, et al., 2005;Sabini, et al., 2003). Ile24, although not being directly involved in any interaction with the amino acids present in the active site of DCK, is present in close proximity to a P-loop residue Gly28 which interacts with an NH_2 residue of Arg128 in the ERS motif.

To determine the functional significance of these 3 coding variants we expressed DCK wild type (WT) and mutant cDNAs (DCK 24mt, DCK 119mt, and DCK 122mt) in BL21 *E. coli* using pET101 expression vectors. The recombinant proteins were obtained 3 hr after IPTG induction and were analyzed for DCK expression by western blotting using a DCK-C-ter antibody. Recombinant DCK wildtype and mutant proteins were expressed in equivalent amounts, whereas no DCK protein was detected with empty pET101 vector (inset in Fig. 2B).

DCK activity was assayed using 12.85 μM ^{3}H CdA and the wildtype and mutant recombinant DCK proteins. As shown in Fig. 2B, the DCK 24mt, DCK 119mt, and DCK 122mt proteins demonstrated 85 ± 5% (13.9 ± 0.8 nmol CdA-MP formed/mg protein · h^{-1}, p = 0.16), 66 ± 2.6% (10.8 ± 0.4 nmol CdA-MP formed/mg protein· hr^{-1}, p = 0.04) and 43 ± 3.7% (7 ± 0.6 nmol CdA-MP formed/mg protein · hr^{-1}, p = 0.04) activity as compared to DCK wildtype protein (16.9 ± 0.05 nmol CdA-MP formed/mg protein · h^{-1}). The empty vector pET101 had undetectable DCK activity.

We further performed substrate kinetic studies using recombinant DCK (WT and mutant) proteins and varying concentrations of ^{3}H-CdA (0.25-12.85 μM). The substrate

JPET Fast Forward. Published on September 12, 2007 as DOI: 10.1124/jpet.107.128595
This article has not been copyedited and formatted. The final version may differ from this version.

velocity curves for the DCK WT, DCK 24mt, DCK 119mt, and DCK 122mt proteins were prepared using the Graphpad Prism software and are shown in Fig. 2C. The apparent Km of DCK-WT was 6.45 ± 0.5 μM. The apparent Km of DCK 24mt (Val24, 9.08 ± 1.3 μM) was approximately 1.5-fold higher than DCK-WT (p=0.058), and the apparent Km of DCK 119mt (Gly119, 3.8 ± 0.7 μM) and DCK 122mt (Ser122, 3.8 ± 0.5 μM) were approximately 1.7-fold lower than Km of the DCK-WT protein (DCK119mt, p=0.007; DCK122mt, p=0.003). The Vmax of DCK 24mt was similar to DCK wildtype (24.76 ± 2.0 vs 25.04 ± 1.0, respectively) while Vmax of DCK 119mt and DCK 122mt were 1.5- and 2.5-fold lower than DCK-WT (15.93 ± 1.1, p=0.0006 and 9.77 ± 0.6, p=0.00009 respectively).

Analysis of DCK activities in HapMap cell lines

DCK activity was assayed in HapMap lymphoblast cell lines homozygous WT for the coding changes (YRI, n = 13) versus HapMap cell lines heterozygous (n = 11) for the coding DCK variants (Ile24Val, YRI, n = 3; Ala119Gly, YRI, n = 4; Pro122Ser, YRI, n = 2; CEPH, n = 3; one YRI HapMap cell line was compound heterozygous for both Ile24Val and Ala119Gly). As shown in Fig. 3, the lymphoblast cell lines from subjects heterozygous for Ile24Val, Ala119Gly, and Pro122Ser demonstrated significantly lower DCK activity as compared to cell lines from homozygous WT subjects (*P* = 0.014, 0.0034, and 0.02, respectively). The DCK activity expressed as pmols CdA-MP formed/mg protein · h^{-1} was 340.8 ± 20, 172.3 ± 47, 159.2 ± 51, and 193.5 ± 57, for genotypes WT/WT, Ile/Val, Ala/Gly, and Pro/Ser, respectively (Fig. 3).

DCK mRNA expression quantification in HapMap cell lines

We determined DCK relative mRNA expression in 87 CEPH (representing ~30 trios with European ancestry) and 90 YRI (30 trios with African ancestry) HapMap samples using the TaqMan gene expression assays from Applied Biosystems.

Fig. 4 shows the range of variability observed in CEPH and YRI ethnic groups with respect to the mRNA expression of DCK. The relative mRNA values shown are $\log_2$ transformed. DCK mRNA expression in the lymphoblast cell lines studied was significantly associated with race (*P* = 0.0001), with subjects with African ancestry

JPET Fast Forward. Published on September 12, 2007 as DOI: 10.1124/jpet.107.128595
This article has not been copyedited and formatted. The final version may differ from this version.

having significantly higher DCK mRNA expression as compared to the subjects with European ancestry.

We further compared DCK mRNA expression and activity in 24 samples (13 homozygous WT and 11 heterozygous for the nonsynonymous coding changes discussed earlier). While the correlation between mRNA and activity was 0.4 in homozygous WT samples, it was as low as to 0.06 in subjects heterozygous for coding variant alleles (P = 0.12). Moreover the homozygous WT genotype (WT/WT) had a significantly greater ratio of DCK activity to DCK mRNA expression than subjects heterozygous for Ala119Gly (P = 0.0008), Pro122Ser (P = 0.02), and Ile24Val (P = 0.04) changes (data not shown).

Association of SNPs with DCK mRNA expression

We also analyzed the association of DCK mRNA expression with SNPs identified in the present study, as well as the ones available from the HapMap database. In the 87 CEPH samples studies, DCK mRNA expression was associated with SNPs in Group1/Block1. For SNPs at *3122* or *35708* positions (representative SNPs in Block1; numbering is with respect to translation start site as +1) the homozygous *3122 CC* (or *35708 TT*) subjects had higher DCK expression compared to the heterozygous subjects (P = 0.02; Fig. 5A and Table 3). None of the other variants were significantly associated with DCK expression among CEPH samples.

Among YRI samples also Group 1/Block 1 SNPs were associated with DCK mRNA expression. Subjects homozygous wildtype for *3122 CC* or *35708 TT* had higher DCK mRNA expression as compared to subjects with at least one *3122T* or *35708C* allele (P = 0.04 for 3122 and P = 0.07 for *35708*; Fig. 5B and Table 3). Interestingly *3122*T (*35708C*) is the major allele in YRI as compared to *3122C* (and 35708T allele) in CEPH samples. Among YRI samples, Group 3Y SNPs (Fig. 5C; represented by *36113* SNP), and a promoter SNP at *-245* (lies in multiple E2F binding sites; Fig. 5D) demonstrated significant association with DCK mRNA expression. In addition to the SNPs mentioned above, the *3547* C>T change representing variants of Group 2Y (and in linkage with Ala119Gly coding change) demonstrated a trend towards higher DCK mRNA levels but was not statistically significant and 29377 C>T change as associated with lower DCK

JPET Fast Forward. Published on September 12, 2007 as DOI: 10.1124/jpet.107.128595
This article has not been copyedited and formatted. The final version may differ from this version.

mRNA levels in YRI samples (p=0.03). The results of these association analyses are summarized in Table 3.

Since none of the intronic variants that were significantly associated with mRNA expression were present at the splice junctions, we screened variants within 100 bp of the splice junction for any changes in splicing enhancer sequences. *In silico* analyses for selected intronic variants at positions 128 (IVS1+36), 28788 (IVS3+53), and 32961 (IVS6+41) were carried out using the web based software ESE finder (http://rulai.cshl.edu/tools/ESE/ESEbkgr.html) which screens for the potential binding affinities for the 4 main serine/arginine-rich (SR) proteins: namely SF2/ASF, SC35, SRp40, and SRp55. SR proteins are a family of structurally-related and highly-conserved splicing factors characterized by 1 or 2 RNA-recognition motifs (RRM) and by a distinctive C-terminal domain highly enriched in RS dipeptides (the RS domain). These proteins bind to splicing enhancer sequences and modulate splicing. The*128 G>C* variant created 1 binding site each for SF2/ASF (score = 1.96, threshold = 1.96), SC35 (score = 3.76, threshold = 2.38), and SRp55 (score = 3.75, threshold = 2.67). The *28778 G>A* and *32961 A>T* had no influence on the splicing enhancer sequences for the binding of SR proteins.

Although additional future molecular biology studies directed towards studying alternate splicing in DCK would provide more insight into functional consequences of this variant, an *in silico* search of various EST/mRNA databases suggests the existence of an alternatively spliced DCK isoform (Clone ID # CD014016) (Jin, et al., 2004) that uses an alternate splice site acceptor sequence in intron 1, and results in an extension of 91 bp at the 3' end of exon 1. The presence/absence of *128 G>C* change could contribute to the formation of this alternatively spliced isoform.

Variability in ara-CTP levels in leukemic blasts from AML patients

As has been published earlier for a subset of patients from the present study (Crews, et al., 2002) in AML patients undergoing ara-C chemotherapy as part of AML97 clinical protocol in St Jude, the intracellular accumulation of ara-CTP was significantly higher when ara-C was given as short daily infusion, as compared to continuous infusion (*P* = 0.014; mean ± SD, 0.56 ± 0.5 vs. 0.33 ± 0.31 nmol ara-CTP/2x10^7 leukemic cells). The

JPET Fast Forward. Published on September 12, 2007 as DOI: 10.1124/jpet.107.128595
This article has not been copyedited and formatted. The final version may differ from this version.

inter-patient variability for blast ara-CTP concentration was 40-fold (n=27; range 0.06-2.43 nmol/2x10^7 cells) when ara-C was administered as short infusion and 101-fold (n=28; range 0.012-1.22 nmol/2x10^7 cells) when given as continuous infusion. The observed variability in the levels of ara-CTP could be due to multiple factors such as expression levels of hENT1, an ara-C uptake transporter, activating enzymes like DCK or inactivating enzymes like cytidine deaminase or 5' nucleotidase etc. Karyotypes, gender, age, and ethnicity had no significant influence on the leukemic ara-CTP levels (P > 0.3 for each of 8 comparisons; data not shown).

Exploratory analysis of association of selective germline DCK SNPs in AML patients

Germline genomic DNA from 55 patients (27 of whom received short daily infusion of ara-C and 28 of whom received continuous infusion of ara-C; 58% White, 22% Black, and 20% with other ethnic backgrounds) was sequenced for exon 3 (as most of the coding variants were observed in exon 3) and exon 7 (including the 3'UTR, as most of the SNPs in this region would capture the major haplotypes in European and African populations). A total of 17 variants were identified including 3 coding changes in exon 3 (*28624 C>T*, Ala100Ala; *28680 C>G*, Ala119Gly; and *28688 C>T*, Pro122Ser), 1 in intron 2, 2 in intron 3, and 11 in the 3'UTR region.

As intracellular ara-CTP concentration differed significantly when ara-C was administered either as short daily infusion or as continuous infusion, we analyzed the data separately within each group. As only 1 patient each had the Ala119Gly and Pro122Ser non-synonymous polymorphisms, we could not perform further analysis. Interestingly, however, the patient heterozygous for Ala119Gly (continuous infusion arm) had the lowest intracellular concentration of ara-CTP (0.012 nmol per 2x10^7 cells; Fig. 6B) within continuous infusion arm. The patient heterozygous for Pro122Ser (short daily infusion arm) had low ara-CTP concentration (0.375 nmol ara-CTP/ 2x10^7 leukemic cells) but was not significantly different from the patients not having the coding change. Both the patients heterozygous for the coding amino acid changes, also relapsed inspite of having favorable cytogenetic abnormalities [inv16 and t(8;21), respectively]. A 3'UTR SNP at position *35708 T>C* (represents Group1/Block 1 in both

European and African Ancestry) was associated with significantly lower intracellular ara-CTP concentrations in patients receiving ara-C as continuous infusion (P = 0.04; Fig. 6A and Table 3).

Discussion:

Ara-C is one of the most effective chemotherapeutic agents used in AML treatment, However resistance of leukemic cells to ara-C and other cytotoxic agents remains a major drawback in the treatment of AML. Several *in vitro* and *in vivo* studies have suggested that the altered activity of enzymes involved in the metabolic activation of ara-C is associated with ara-C response (Colly, et al., 1987; Hubeek, et al., 2006; Kobayashi, et al., 1994; Owens, et al., 1992). DCK is a rate limiting enzyme involved in the activation of ara-C to ara-CTP.

Because of the critical role of DCK in the activation of ara-C, we sought to determine whether inherent genetic variation in the DCK gene could account for the inter-patient variability observed in ara-CTP accumulation. We identified 3 coding variants (Ile24Val, Ala119Gly and Pro122Ser) in DCK with altered activity. Compared to DCK-wild type (WT) protein, the activity of the recombinant DCK-Ile24Val, DCK-Ala119Gly, and DCK-Pro122Ser proteins was 85±5%, 66±3%, and 43±4%, respectively (Fig. 2). The DCK coding mutants were also associated with altered kinetics. The kinetic data presented here indicates that DCK-24mt might have slightly decreased (but not significant) substrate affinity as compared to DCK-WT. Further, DCK-119mt and 122mt have significantly lower apparent Km and Vmax as compared to DCK-WT, we speculate that these mutants although have increased substrate affinity would have compromised ability to catalyze the formation of the product. Thus these variants may have a slower turnaround time and will trap the substrate for longer time. Whether these mutants have altered binding affinities for other nucleoside analogs remains to be determined. These results suggest that a subset of population might have different rate of drug activation which might influence the response.

HapMap cell lines heterozygous for these coding changes were associated with a significant reduction in DCK activity as compared to cell lines that were homozygous WT (Fig. 3). The observed low levels of DCK activity in subjects that are heterozygous for DCK coding changes, in spite of the fact that they have one WT allele present, may be due to several reasons: 1) since DCK acts as a dimer it is possible that dimerization of mutant-DCK with WT-DCK could have an influence DCK activity; 2) the coding change at position *28688* (Pro122Ser) occurs on the same haplotype as Block 1 SNPs,

which are associated with DCK mRNA expression, and it is possible that these SNPs are having a synergistic effect thus resulting in an affect on both mRNA expression as well as DCK activity; 3) Ala119Gly and Pro122Ser coding mutants are present in close proximity to the ERS in DCK active site motif that interacts with the substrate. As both these mutants have higher substrate affinity but lower activity, it is possible that the mutant protein forms might be trapping the substrate, making less of it available for the WT-DCK protein; and 4) it has been documented in literature that DCK could be inactivated by phosphorylation of Ser residues (Smal, et al., 2006). Whether Ser122 is a target of phosphorylation and DCK inactivation is not known at the present time, but will be determined through future studies. 5) The half life of DCK mRNA and/or protein could be influenced by the variants; it is possible that steady state levels of DCK protein (dependent or independent of coding changes) might be contributing to the observed differences in lymphoblast cell lines vs. recombinant protein. Although all the cell lines are grown in the same environment and have comparable proliferation rate, the differential effects of endogenous DCK substrates on its activity could not be completely ruled out.

In the translational part of our study, we genotyped selected DCK SNPs in AML patient samples and analyzed their association with intracellular concentrations of ara-CTP in leukemic blasts.

We observed significantly higher DCK mRNA expression in subjects with African ancestry (YRI) as compared to European ancestry (CEPH). Further SNPs in haplotype Block 1 were associated with DCK mRNA expression in both European (CEPH) and African (YRI) ethnic groups (Fig. 5A-B). Within AML patients, the C allele of 3'UTR SNP at position *35708* was associated with lower blast ara-CTP concentrations when ara-C was given as continuous infusion (Fig. 6). This observation was further supported by the association of *35708 C* allele with lower expression of DCK mRNA in HAPMAP cell lines, suggesting that this allele might be responsible for lower ara-CTP formation in AML patients. Although a similar association of *35708* SNP was not observed for patients receiving ara-C as short daily infusion; one of the reasons might be the lesser number of patients (n = 5) carrying the mutant allele in the short infusion arm as compared to the continuous infusion arm (n = 9). It is also possible that short daily

infusion of ara-C produces high peak concentrations that could thus overcome the limitation produced by the DCK mutation.

The promoter variants (represented by *-198* and *-357* in the present study) identified previously (Shi, et al., 2004), occurred with low frequency (0.03 as compared to 0.15 in the Chinese population) and showed no association with DCK mRNA expression in European (CEPH) samples, these variants were not found in YRI samples. The K242P amino acid change reported earlier (Joerger, et al., 2006) was not observed in CEPH or YRI samples.

Although we used pharmacokinetic data from the St. Jude AML97 protocol to analyze the association of the genetic variants in DCK with blast ara-CTP levels, this study was not designed to evaluate the pharmacogenetics of ara-C. Thus the results of the exploratory analysis in AML patients need to be confirmed in a larger patient population. We also acknowledge the fact that variability in the intracellular levels of ara-CTP is regulated by multiple enzymes in ara-C activation pathway with DCK being one of the candidates. In order to achieve a full understanding of the genetic basis for the variability observed in intracellular concentration of ara-CTP, future studies in our lab are also directed towards other enzymes of relevance in the metabolic pathway of ara-C. The results of the present study will be confirmed in the ongoing St. Jude AML02 protocol which is a larger clinical study aimed at enrolling approximately 200 de novo AML patients.

Once confirmed in a larger patient population, we anticipate that genetic variants in DCK and other genes in the ara-C metabolic pathway could help, in part, to predict the intracellular levels of the active metabolite, ara-CTP and hence responsiveness to ara-C. Although AML is a very heterogeneous disease with different subtypes that are of prognostic significance, the pharmacogenetics of ara-C could help in better understanding of drug responsiveness and guide us to develop individualized chemotherapy in cancer patients receiving nucleoside analogs.

In summary, we have identified novel coding, promoter, intronic, and 3'-UTR genetic variants at the DCK locus in 2 major ethnic groups. Three coding variants and SNPs in Block 1 are associated with DCK activity and expression, respectively. Further we observed ethnic differences in DCK mRNA expression in subjects with European or

African ancestry. Furthermore, we have shown in a pilot study the clinical implication of DCK polymorphisms on the intracellular concentrations of the ara-CTP in leukemic blasts of AML patients undergoing treatment with ara-C.

Acknowledgements:

We are grateful to Xueyuan Cao for assistance with statistical analysis, Hartwell Center for Bioinformatics for sequencing and for running Agilent arrays to determine the quality of RNA samples and the Tissue Resource Committee at St. Jude Children's Research Hospital for providing DNA from AML patients. Dr. Donald Samulack for providing the editorial assistance, Dr. David Finkelstein for assistance in preparation of graphs and PAAR cell core for providing the DNA and RNA samples on HapMap cell lines.

References

Arner ES, Spasokoukotskaja T and Eriksson S (1992) Selective assays for thymidine kinase 1 and 2 and deoxycytidine kinase and their activities in extracts from human cells and tissues. *Biochem Biophys Res Commun* **188**:712-718.

Crews KR, Gandhi V, Srivastava DK, Razzouk BI, Tong X, Behm FG, Plunkett W, Raimondi SC, Pui CH, Rubnitz JE, Stewart CF and Ribeiro RC (2002) Interim comparison of a continuous infusion versus a short daily infusion of cytarabine given in combination with cladribine for pediatric acute myeloid leukemia. *J Clin Oncol* **20**:4217-4224.

Cros E, Jordheim L, Dumontet C and Galmarini CM (2005) Problems related to resistance to cytarabine in acute myeloid leukemia. *Leuk Lymphoma* **45**:1123-1132.

Galmarini CM, Thomas X, Calvo F, Rousselot P, Jafaari AE, Cros E and Dumontet C (2002) Potential mechanisms of resistance to cytarabine in AML patients. *Leuk Res* **26**:621-629.

Jin P, Fu GK, Wilson AD, Yang J, Chien D, Hawkins PR, Au-Young J and Stuve LL (2004) PCR isolation and cloning of novel splice variant mRNAs from known drug target genes. *Genomics* **83**:566-571.

Joerger M, Bosch TM, Doodeman VD, Beijnen JH, Smits PH and Schellens JH (2006) Novel deoxycytidine kinase gene polymorphisms: a population screening study in Caucasian healthy volunteers. *Eur J Clin Pharmacol* **62**:681-684.

Johnsamuel J, Eriksson S, Oliveira M and Tjarks W (2005) Docking simulation with a purine nucleoside specific homology model of deoxycytidine kinase, a target enzyme for anticancer and antiviral therapy. *Bioorg Med Chem* **13**:4160-4167.

Kakihara T, Fukuda T, Tanaka A, Emura I, Kishi K, Asami K and Uchiyama M (1998) Expression of deoxycytidine kinase (dCK) gene in leukemic cells in childhood: decreased expression of dCK gene in relapsed leukemia. *Leuk Lymphoma* **31**:405-409.

Kroep JR, Loves WJ, van der Wilt CL, Alvarez E, Talianidis I, Boven E, Braakhuis BJ, Van Groeningen CJ, Pinedo HM and Peters GJ (2002) Pretreatment deoxycytidine kinase levels predict in vivo gemcitabine sensitivity. *Mol Cancer Ther* **1**:371-376.

Kufe DW, Major PP, Egan EM and Beardsley GP (1980) Correlation of cytotoxicity with incorporation of ara-C into DNA. *J Biol Chem* **255**:8997-900.

Kufe DW, Munroe D, Herrick D, Egan E and Spriggs D (1984) Effects of 1-beta-D-arabinofuranosylcytosine incorporation on eukaryotic DNA template function. *Mol Pharmacol* **26**:128-134.

Major PP, Egan EM, Beardsley GP, Minden MD and Kufe DW (1981) Lethality of human myeloblasts correlates with the incorporation of arabinofuranosylcytosine into DNA. *Proc Natl Acad Sci U S A* **78**:3235-3239.

Nickerson DA, Tobe VO and Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* **25**:2745-2751.

Raza A, Gezer S, Anderson J, Lykins J, Bennett J, Browman G, Goldberg J, Larson R, Vogler R and Preisler HD (1992) Relationship of [3H]Ara-C incorporation and response to therapy with high-dose Ara-C in AML patients: a Leukemia Intergroup study. *Exp Hematol* **20**:1194-1200.

Sabini E, Ort S, Monnerjahn C, Konrad M and Lavie A (2003) Structure of human dCK suggests strategies to improve anticancer and antiviral therapy. *Nat Struct Biol* **10**:513-519.

Sarkar M, Han T, Damaraju V, Carpenter P, Cass CE and Agarwal RP (2005) Cytosine arabinoside affects multiple cellular factors and induces drug resistance in human lymphoid cells. *Biochem Pharmacol* **70**:426-432.

Shi JY, Shi ZZ, Zhang SJ, Zhu YM, Gu BW, Li G, Bai XT, Gao XD, Hu J, Jin W, Huang W, Chen Z and Chen SJ (2004a) Association between single nucleotide polymorphisms in deoxycytidine kinase and treatment response among acute myeloid leukaemia patients. *Pharmacogenetics* **14**:759-768.

van der Wilt CL, Kroep JR, Loves WJ, Rots MG, Van Groeningen CJ, Kaspers GJ and Peters GJ (2003a) Expression of deoxycytidine kinase in leukaemic cells compared with solid tumour cell lines, liver metastases and normal liver. *Eur J Cancer* **39**:691-697.

Wang JJ, Selawry OS, Vietti TJ and Bodey GP, Sr. (1970) Prolonged infusion of arabinosyl cytosine in childhood leukemia. *Cancer* **25**:1-6.

Wolfinger R.D. (1996) Heterogeneous Variance: Covariance Structures for Repeated Measures. *Journal of Agricultural, Biological, and Environmental Statistics* **1**:205-230.

Foot Notes:

This work is supported in part by the NIH/NIGMS Pharmacogenetics Research Network and Database (U01GM61374, http://pharmgkb.org) under grant U01 GM61393, in part by the cancer center support grant (CA-21765) and by the American Lebanese Syrian Associated Charities (ALSAC). Jessica Gresham was supported in part by 5 R25 CA23944 from the National Cancer Institute.

JPET Fast Forward. Published on September 12, 2007 as DOI: 10.1124/jpet.107.128595
This article has not been copyedited and formatted. The final version may differ from this version.

Legends for Figures:

Fig. 1. A). Genetic variant MAP of DCK gene. Genetic variants were identified in the present study and from the HapMap database are shown and numbered with respect to translational start site being +1. Variants from HapMap are depicted by taller arrows and those overlapping with the present study are represented by taller arrows and an asterisk after the nucleotide change. Variants in the promoter are indicated by a negative number and an asterisk before the nucleotide change (shorter arrows). Coding changes are boxed and the associated amino acid change is indicated. YRI* indicates that the coding change was identified only in the African ethnic group. **B) LD plot of DCK in samples with European (CEPH) and African (YRI) ancestry.** LD plots were generated in Haploview using genotype data from the present study and from HapMap in both CEPH and YRI samples. The color scheme is: white when $r^2 = 0$; shades of grey light $0<r^2<1$; and black when $r^2 = 1$. SNPs that are linked $r^2 > 0.8$ (and picked by tagger program) are categorized in the same groups. Group 1, represented by Block 1 in the LD plot, has common SNPs in both ethnic groups. Group number ends in C for CEPH and Y for YRI, and group frequencies indicated in column 2.

Fig. 2. A) Conservation of DCK coding amino acid changes in 5 different species. The DCK amino acid sequences for cow, human, monkey (macaque), mouse, and rat were obtained from NCBI. Vector NTI was used to align the sequences. Amino acid polymorphisms at positions 24, 119, and 122 are boxed and the variant change is indicted on top. The ERS motif in the active site of DCK and Gly28 is boxed by a dotted line. The numbering of the amino acid residues is indicated on the top. **B) Activity of recombinant human DCK WT and mutant proteins.** The columns represent DCK activity determined using ^{3}H-CdA as a substrate. DCK activity was measured as nmol CdA-MP formed/mg protein $\cdot$ hr^{-1}. Activity of DCK-24mt, 119mt, and 122mt proteins is compared with DCK-WT protein. Inset shows a representative western blot when equal amount of total protein was loaded for WT and mutant DCK proteins; all of the recombinant DCK variants were expressed at equivalent levels. **C) Substrate kinetics of recombinant wildtype (WT) and mutant (mt) DCK.** DCK activity was assayed

using ^{3}H-CdA as a substrate and by measuring CdA-MP formed/mg protein · hr^{-1}. Substrate kinetics was estimated at least 3 times. Each data point represents the mean with error bars from an experiment. Graphpad prism was used to determine the kinetics.

Fig. 3. DCK activity in lymphoblast cell lines that are homozygous WT vs. heterozygous for coding variants. The genotype as well as numbers of samples analyzed are indicated on the X axis and DCK activity measured as pmol CdA-MP formed/mg protein · hr^{-1} is represented on the Y axis. # indicates 1 patient was compound heterozygous for both Ile24Val and Pro122Ser polymorphisms. * indicates that this group includes 3 CEPH cell lines in addition to YRI cell lines. All other cell lines shown with coding changes are YRI. **p<0.01, *p<0.05.

Fig. 4. DCK mRNA expression determined by real-time PCR quantitation (Taqman) in RNA from EBV- transformed lymphoblast cell lines derived from subjects with European (CEPH) and African (YRI) ancestry. Median values for Log_2DCK mRNA levels are indicated by a horizontal line for each ethnic group. ***p<0.001.

Fig. 5. Association of DCK mRNA expression in lymphoblastoid cell lines with genetic variants at DCK locus. Box plots for the association of *3122 C>T* SNP with DCK mRNA levels (representing LD Block1 and in LD with 35708) in CEPH and YRI samples (5A and 5B). Box plots representing the association of *36113 C>T* SNP (Group 3Y), and a promoter polymorphism at *-245 C>G*, with DCK mRNA expression in YRI samples are shown in 5C and 5D, respectively. Plots show medians as a line between boxes which represent 1st and the 3rd quartiles; the whiskers represent the range after excluding the outliers. The outliers are defined as data points which fall outside of the 1st and 3rd quartiles by more than 1.5 times the inter-quartile range. Circles falling outside the whiskers represent outliers. * p<0.05.

Fig. 6. Association of blast ara-CTP levels with DCK polymorphisms. Box plots showing association of Block 1 SNP (*35708 C>T*) with leukemic blast ara-CTP concentrations in AML patients receiving ara-C as continuous infusion.

JPET Fast Forward. Published on September 12, 2007 as DOI: 10.1124/jpet.107.128595
This article has not been copyedited and formatted. The final version may differ from this version.

Table.1 PCR primer sequences and cycle conditions for re-sequencing of DCK.

Primers	Sequence 5' to 3'	Cycle conditions	Size of the amplicon
DCK5'UTR-F DCk5'UTR-R	AGCGAACCAAGTGCTTCAAG GGGAATCGCACTGCCCACT	59°C for 1min; 5% DMSO; 1.5 mM MgCl2	1kb
DCKex1-F DCKex1-R	AGTGGGCAGTGCGATTCCCA GGTAAGGGAAGGATGCTCTG	59°C for 1min; 5% DMSO; 1.5 mM MgCl2	0.93kb
DCKex2-F DCKex2-R	ATTGGGCAGGGAGCCTTTTCA TGGTTGAATCCATATCAATCTG	55°C for 40 sec; 1.5 mM $MgCl_2$	0.5kb
DCKex3.4-F DCKex3.4-R	AAGATCTAAGGATTTTCCAGAC AACTTGCCCAGTCTGGGATT	55°C for 2 min; 1.5 mM $MgCl_2$	1.5kb
DCKex5.6-F DCKex5.6-R	AGTACTGCTTGGCTTAGAGC TAGCTCAGGCCACAAAGCTAG	55°C for 2 min; 1.5 mM $MgCl_2$	1.5kb
DCKex7.3'UTR-F DCKex7.3'UTR-R	ATGCAATGGCATTGTGGTAGT GCATAGCAGAAGAGAAAATTCT	57°C for 2 min; 2 mM $MgCl_2$	1.8kb
DCK-Intron1-F DCK-Intron1-R	CCAGAGCATCCTTCCCTTAC TGAAAAGGCTCCCTGCCCAAT	59°C for 1 min; 1.5 mM $MgCl_2$	4 kb

JPET Fast Forward. Published on September 12, 2007 as DOI: 10.1124/jpet.107.128595
This article has not been copyedited and formatted. The final version may differ from this version.

Table. 2 Frequencies of genetic variants identified in DCK in the present study.

Position with respect to translation start site as +1 (rs number)	SNP/indel	Location	Coding SNPs	Minor Allele Frequency		Observed heterozygosity	
				CEPH	YRI	CEPH	YRI
-1372	G>A	5'UTR		0.000	0.009	0	0.011
-1351	A>G	5'UTR		0.000	0.042	0	0.1
-1321	C>T	5'UTR		0.067	0.025	0.124	0.033
-815	G>A	5'UTR		0.008	0.000	0.011	0
-435	C>T	5'UTR		0.000	0.017	0	0.022
-357[#]	C>G	5'UTR		0.025	0.000	0.034	0
-245	G>C	5'UTR		0.000	0.033	0	0.056
-240	G>T	5'UTR		0.008	0.000	0.023	0
-198[#]	C>T	5'UTR		0.025	0.000	0.034	0
-139	C>A	5'UTR		0.000	0.008	0	0.022
-52	G>A	5'UTR		0.000	0.025	0	0.044
-33	C>A	5'UTR		0.000	0.008	0	0.011
70	A>G	Exon 1	Ile24Val	0.000	0.025	0	0.044
121	A>G	Intron 1		0.008	0.000	0.011	0
128	G>C	Intron 1		0.025	0.000	0.067	0
160	C>A	Intron 1		0.000	0.008	0	0.011
325	G>C	Intron 1		0.009	0.000	0.012	0
759	A>G	Intron 1		0.025	0.15	0.047	0.29
1090	T>A	Intron 1		0.000	0.026	0	0.068
1115	A>T	Intron 1		0.000	0.073	0	0.133
1123	A>T	Intron 1		0.000	0.017	0	0.045
1124	T>A	Intron 1		0.009	0.009	0.013	0.025
1201	C>T	Intron 1		0.000	0.075	0	0.138
1202(rs2035576)	T>G	Intron 1		0.000	0.192	0	0.367
1333	A>G	Intron 1		0.000	0.018	0	0.024
1718	T>C	Intron 1		0.000	0.065	0	0.122
1901	T>C	Intron 1		0.017	0.000	0.033	0
1948	A>G	Intron 1		0.000	0.017	0	0.045
2162	G>A	Intron 1		0.000	0.026	0	0.057
3051	T>A	Intron 1		0.000	0.017	0	0.045
3122(rs3775289)	C>T	Intron 1		0.050	0.767	0.111	0.33
3192	C>T	Intron 1		0.000	0.010	0	0.026
3547	C>T	Intron 1		0.000	0.017	0	0.045
3558	T>C	Intron 1		0.026	0.275	0.037	0.372
3644	T>C	Intron 1		0.184	0.000	0.228	0
4058(rs9993633)	G>A	Intron 1		0.050	0.350	0.111	0.444
4122(rs7693891)	T>G	Intron 2		0.000	0.033	0	0.056
4369	'tagataaaag' deletion	Intron 2		0.000	0.125	0	0.211
4449	C>T	Intron 2		0.000	0.025	0	0.045
4461(rs6446988)	G>A	Intron 2		0.050	0.767	0.111	0.33
28624	C>T	Exon 3	Ala100Ala	0.042	0.000	0.056	0
28680	C>G	Exon 3	Ala119Gly	0.000	0.017	0	0.045
28688	C>T	Exon 3	Pro122Ser	0.025	0.017	0.045	0.022

28778	G>A	Intron 3	0.025	0.000	0.056	0
29377	C>T	Intron 3	0.000	0.009	0	0.012
29507	A>T	Intron 3	0.000	0.008	0	0.022
29559	A>C	Intron 3	0.000	0.017	0	0.045
31942	C>T	Intron 4	0.000	0.025	0	0.034
32421	G>A	Intron 5	0.009	0.000	0.015	0
32484(rs936869)	C>T	Intron 5	0.050	0.767	0.111	0.33
35708(rs4643786)	T>C	3'UTR	0.050	0.767	0.111	0.33
35874	T>G	3'UTR	0.000	0.010	0	0.025
36088	C>A	3'UTR	0.000	0.026	0	0.057
36113	C>T	3'UTR	0.000	0.026	0	0.057
36116	T>C	3'UTR	0.000	0.115	0	0.222
36283	A>G	3'UTR	0.000	0.017	0	0.035
36498	G>A	3'UTR	0.000	0.035	0	0.047
36668	C>T	3'UTR	0.000	0.091	0	0.181
36737	C>G	3'UTR	0.008	0.000	0.011	0
36791	G>A	3'UTR	0.033	0.000	0.056	0
36865(rs16845677)	A>T	3'UTR	0.000	0.127	0	0.213
36969	A>G	3'UTR	0.000	0.032	0	0.067
37017	'tta'deletion	3'UTR	0.000	0.123	0	0.235
37179(rs4525938)	T>A	3'UTR	0.050	0.66	0.19	0.48

All the SNPs were in accordance with Hardy Weinberg Equilibrium.
rs number for SNPs identified in the present study and shared with HAPMAP project.
[#] SNPs referred as -360 and -201 by Shi et al., 2004

Table 3. Association of DCK SNPs with DCK mRNA expression in lymphoblast cell lines from subjects with European (CEPH) and African (YRI) ancestry.

Polymorphism	Phenotype I: DCK mRNA expression				Phenotype II: Ara-CTP nmol /2x10^7 leukemic cells	
	CEPH (n=85)		YRI (n=90)		AML patients receiving Ara-C-Continuous Infusion Arm	
	Genotype	mean±SD	Genotype	mean±SD	Genotype	mean±SD
Group 1/Block1	CC (n=75)	84±37	CC (n=7)	163±27		
represented by 3122 C>T	CT (n=10)	57±18	CT (n=27)	134±44		
			TT (n=56)	130±39		
	CC vs. CT, *p=0.02		CC vs. CT+TT, *p=0.04			
35708 T>C	TT (n=75)	84±37	TT (n=6)	155±44	TT (n=18)	0.48±0.35
	CT (n=10)	57±18	CT (n=27)	135±41	CC+CT (n=9)	0.22±0.17
			CC (n=57)	130±18		
	TT vs. CT, *p=0.02		TT vs. CT+CC, p=0.07		TT vs. CC+CT, *p=0.04$^{\$}$	
Group2Y			CC (n=80)	136±41		
represented by 3547C>T and linked to Ala119Gly			CT (n=3)	170±47		
			CC vs. CT, p=0.16 (0.08$^{\#}$)			
Group3Y			CC (n=75)	141±41		
represented by 36113C>T			CT (n=5)	97±22		
			CC vs. CT, *p=0.01			
-245 G>C			GG (n=85)	138±42		
			GC (n=5)	95±26		
			GG vs. GC, *p=0.01			
29377 C>T			CC (n=86)	138±41		
			CT (n=3)	85±6		
			CC vs. CT, *p=0.03			

$^{\#}$ p value when the analysis was performed in 60 independent parents; $^{\$}$ Excludes the sample heterozygous for Ala119Gly
Genotype data was not available for Group2Y, Group3Y and 29377 SNPs in 7, 10 and 1 YRI samples, respectively. * p<0.05

JPET Fast Forward. Published on September 12, 2007 as DOI: 10.1124/jpet.107.128595
This article has not been copyedited and formatted. The final version may differ from this version.

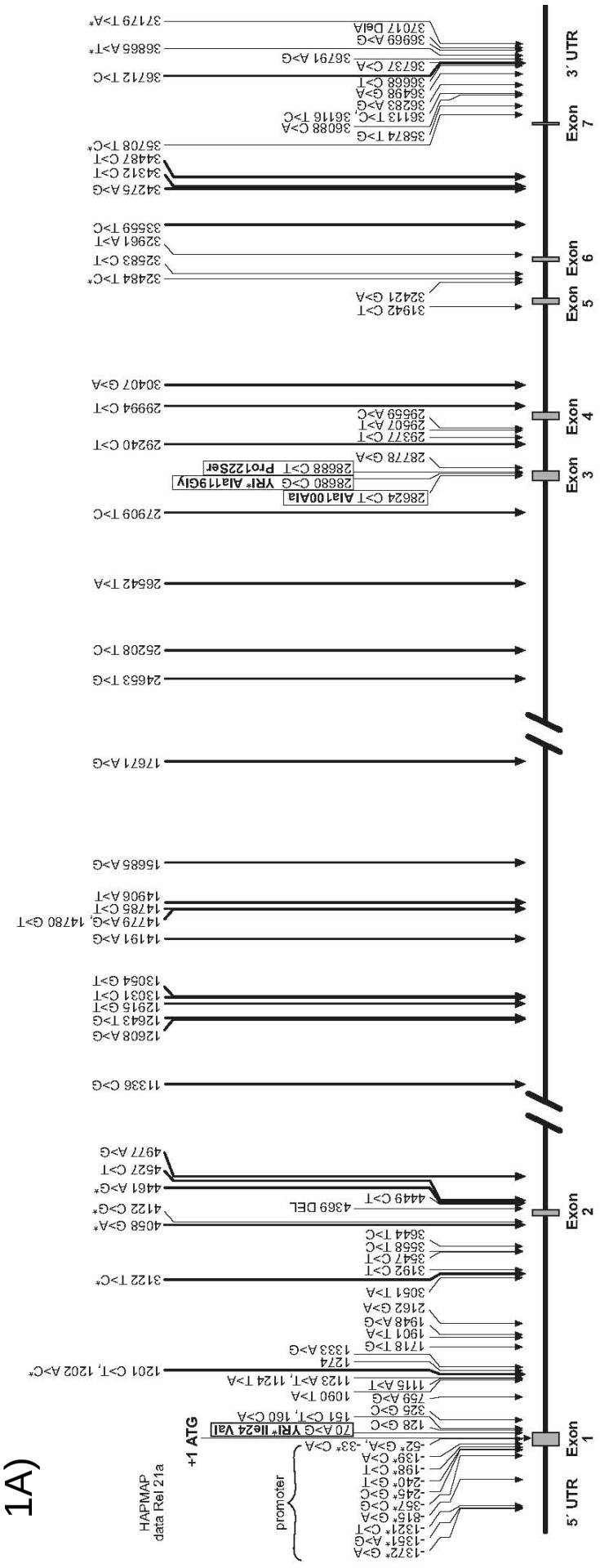
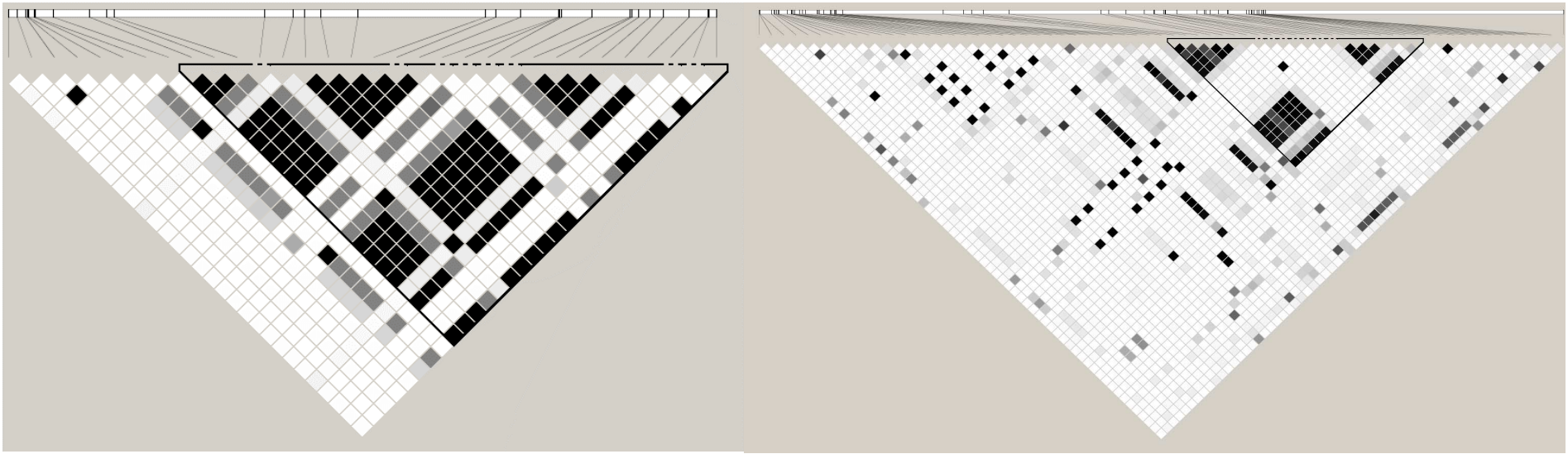


Figure 1A

1B) CEPH (European Ancestry)

YRI (African Ancestry)



CEPH

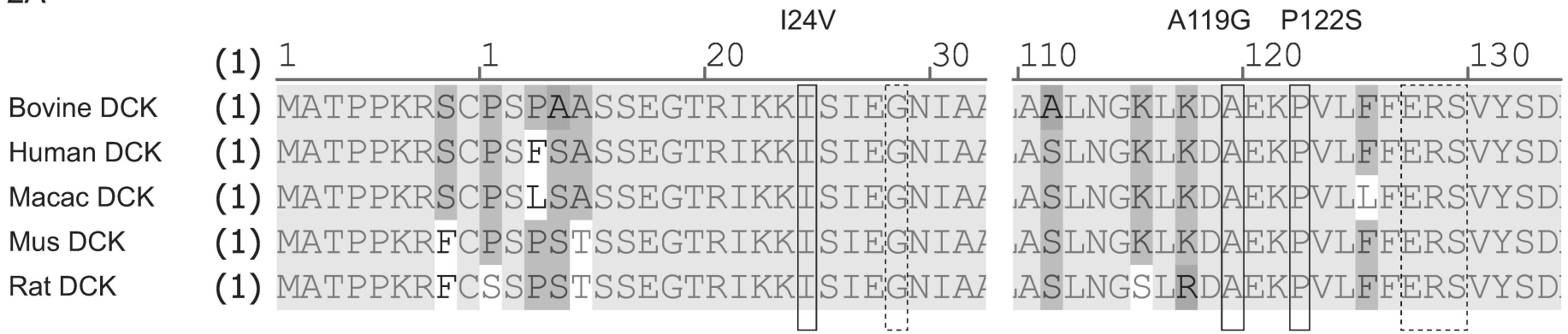
Group	Frequency	SNPs
Group1/ Block1	0.05	3122, 4058, 4461, 14779, 15685, 17671, 24653, 25208, 26542, 32484, 32583, 32961, 33559, 35708, 37179
Group2C	0.025	128, 12643, 30407
Group3C	0.025	-198, -357
Group4C	0.45	14191, 34312

YRI

Group	Frequency	SNPs
Group1/ Block1	0.767	3122, 4461, 14779, 15685, 24653, 25208, 26542, 32484 32583, 32961, 33559, 35708
Group2Y	0.017	1090, 1123, 1948, 3051, 3547, 28680, 29559
Group3Y	0.026	2162, 36088/36113
Group4Y	0.01	-139, 1124
Group5Y	0.01	3192, 29507
Group6Y	0.19	1202, 34275
Group7Y	0.70	17671, 37179
Group8Y	0.11	36116, 37017
Group9Y	0.31	4977, 27909

Figure1B

2A



2B

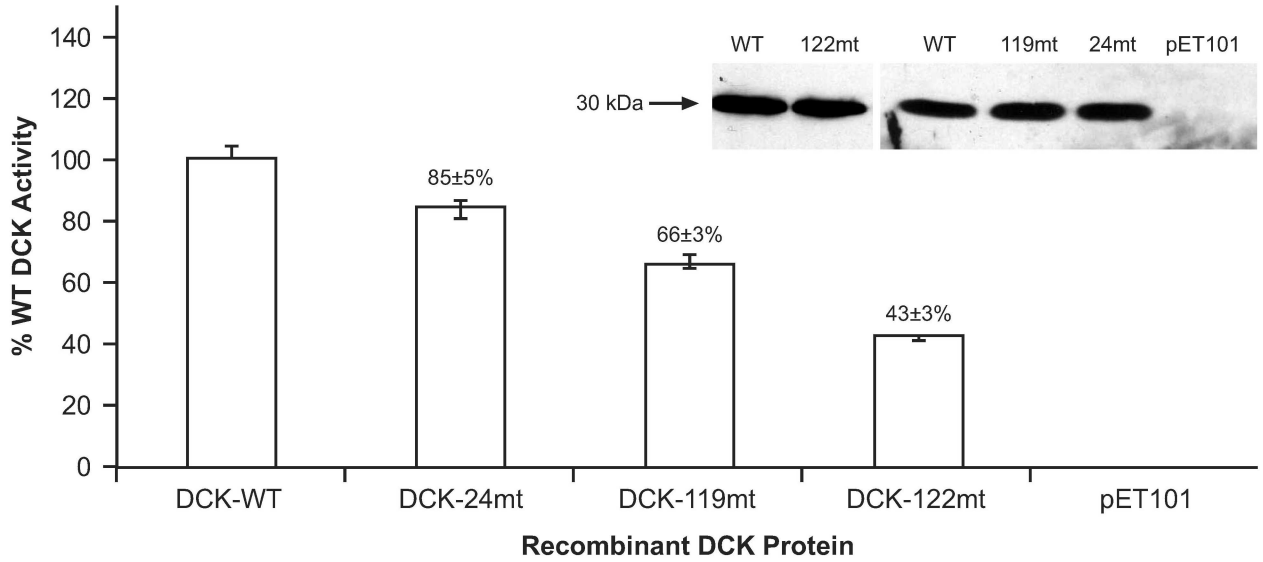


Figure 2A and 2B

2C

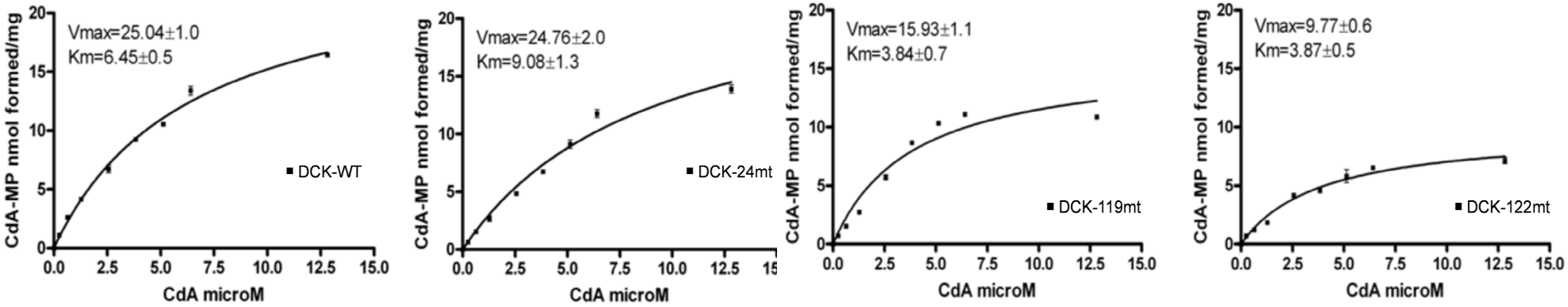


Figure 2C

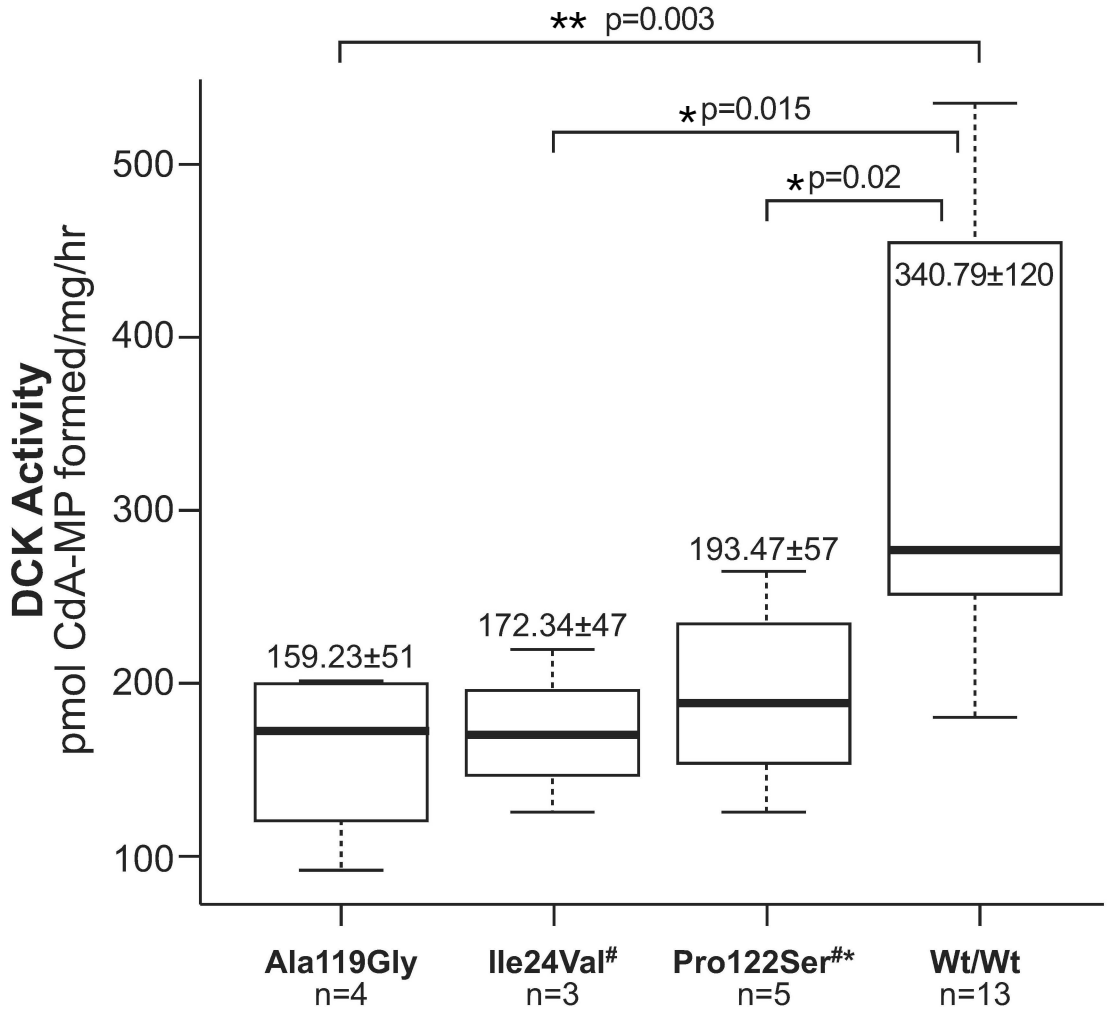
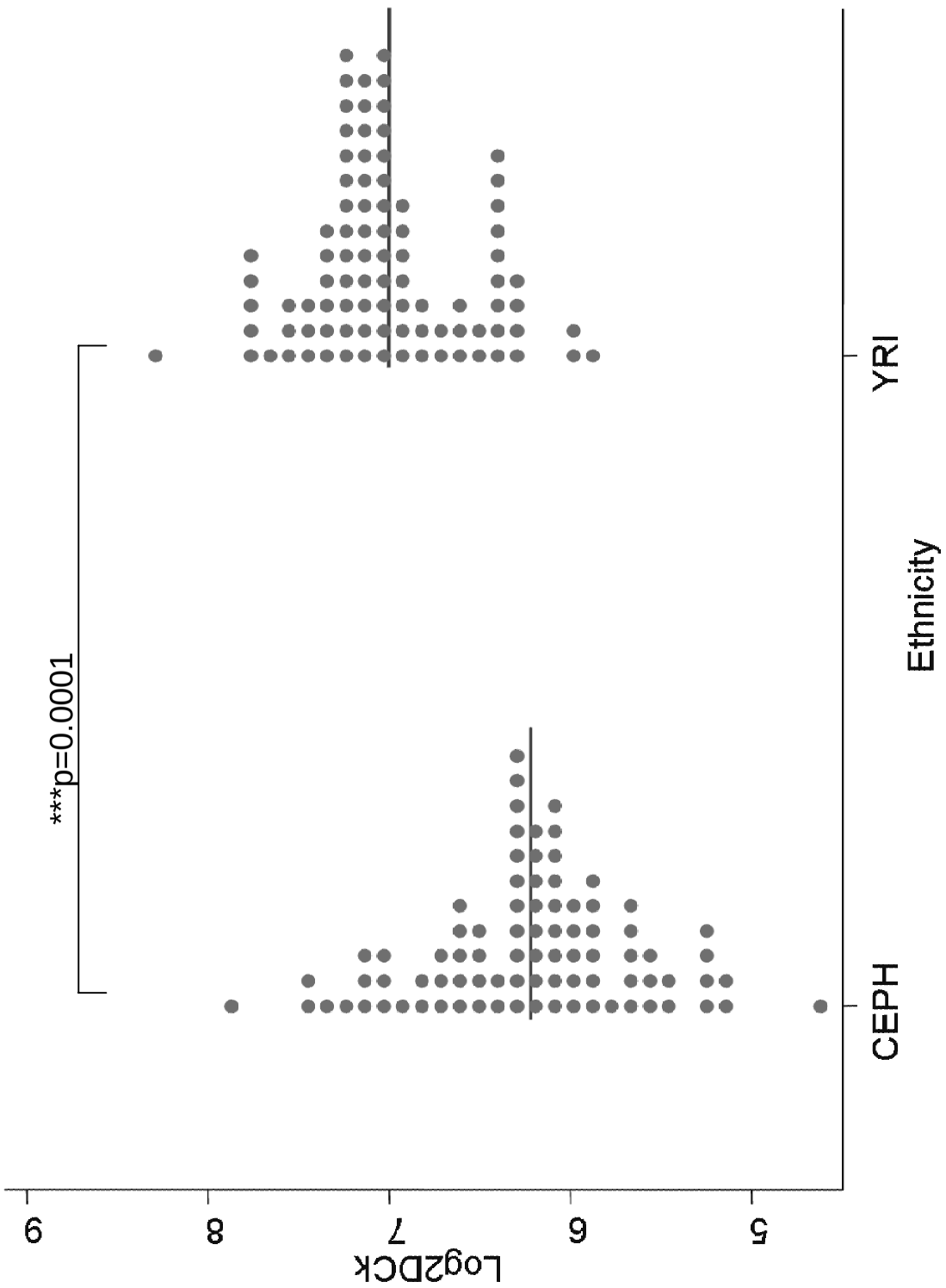


Figure 3

JPET Fast Forward. Published on September 12, 2007 as DOI: 10.1124/jpet.107.128595
This article has not been copyedited and formatted. The final version may differ from this version.

Figure 4



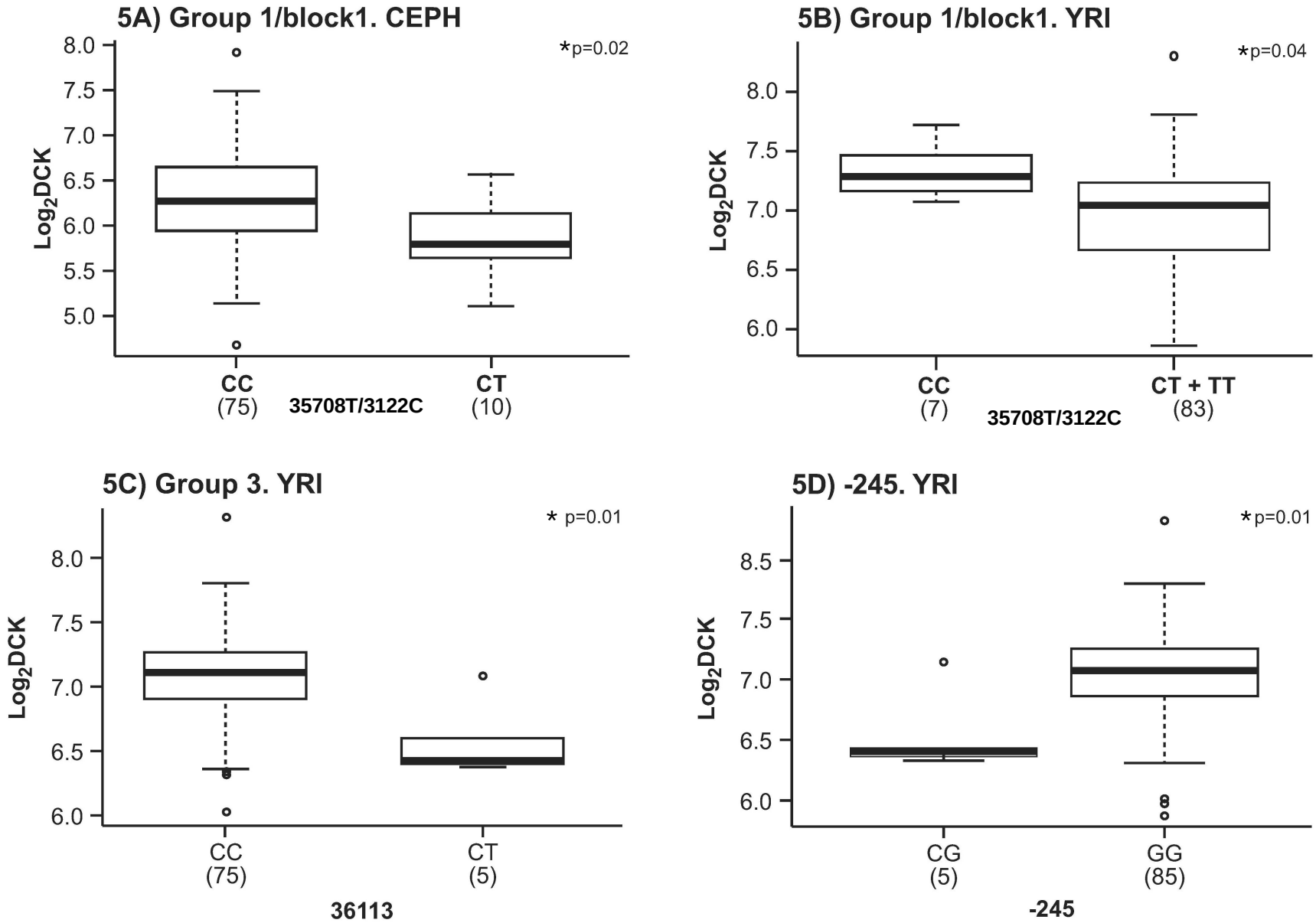


Figure 5

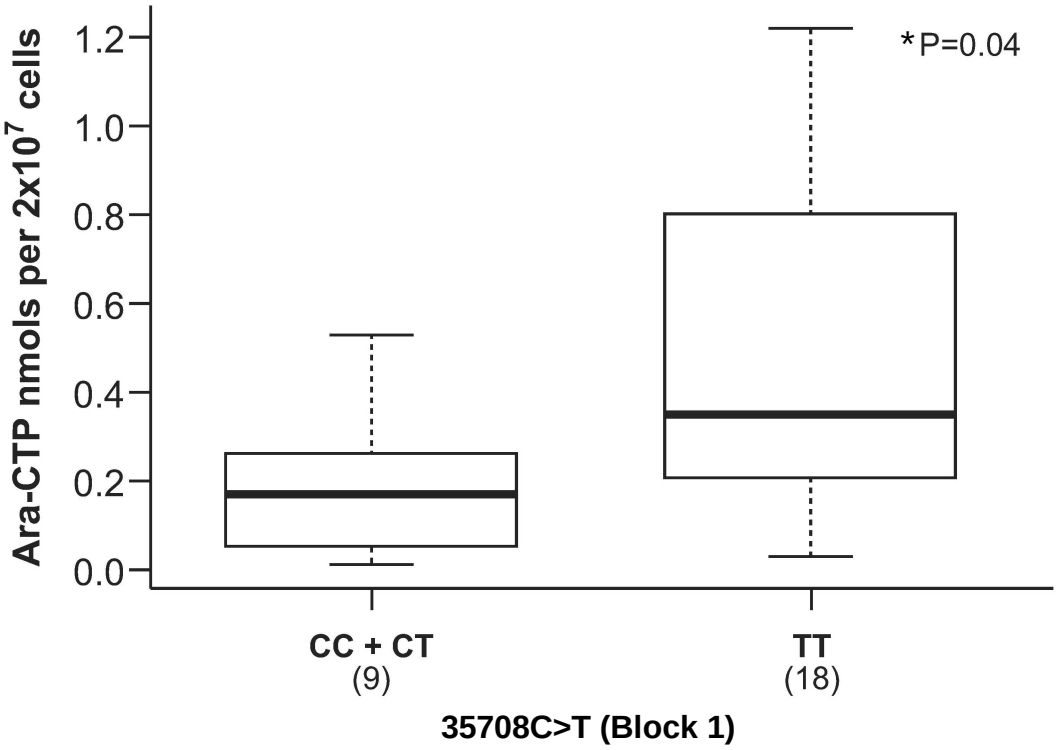


Figure 6