DOI:10.1158/1078-0432.CCR-05-1969

Clinical Cancer Research

Gemcitabine Pharmacogenomics: Cytidine Deaminase and Deoxycytidylate Deaminase Gene Resequencing and Functional Genomics

Judith A. Gilbert, Oreste E. Salavaggione, Yuan Ji, et al.

Clin Cancer Res 2006;12:1794-1803. Published online March 21, 2006.

DOI:10.1158/1078-0432.CCR-05-1969

Gemcitabine Pharmacogenomics: Cytidine Deaminase and Deoxycytidylate Deaminase Gene Resequencing and Functional Genomics

Judith A. Gilbert,[1] Oreste E. Salavaggione,[1] Yuan Ji,[1] Linda L. Pelleymounter,[1] Bruce W. Eckloff,[2] Eric D. Wieben,[2] Matthew M. Ames,[1] and Richard M. Weinshilboum[1]

Abstract

Purpose: Gemcitabine is a nucleoside analogue with activity against solid tumors. Gemcitabine metabolic inactivation is catalyzed by cytidine deaminase (CDA) or, after phosphorylation, by deoxycytidylate deaminase (DCTD). We set out to study the pharmacogenomics of CDA and DCTD.

Experimental Design: The genes encoding CDA and DCTD were resequenced using DNA from 60 African American and 60 Caucasian American subjects. Expression constructs were created for nonsynonymous coding single nucleotide polymorphisms (cSNP) and reporter gene constructs were created for 5′-flanking region polymorphisms. Functional genomic studies were then conducted after the transfection of mammalian cells.

Results: *CDA* resequencing revealed 17 polymorphisms, including one common nonsynonymous cSNP, 79 A>C (Lys27Gln). Recombinant Gln27 CDA had 66 ± 5.1% (mean ± SE) of the wild-type (WT) activity for gemcitabine but without a significant decrease in level of immunoreactive protein. The apparent K_m (397 ± 40 μmol/L) for the Gln27 allozyme was significantly higher than that for the WT (289 ± 20 μmol/L; $P < 0.025$). *CDA* 5′-flanking region reporter gene studies showed significant differences among 5′-flanking region haplotypes in their ability to drive transcription. There were 29 SNPs in *DCTD*, including one nonsynonymous cSNP, 172 A>G (Asn58Asp), in Caucasian American DNA. Recombinant Asp58 DCTD had 11 ± 1.4% of WT activity for gemcitabine monophosphate with a significantly elevated level of immunoreactive protein. No *DCTD* polymorphisms were observed in the initial 500 bp of the 5′-flanking region.

Conclusions: These results suggest that pharmacogenomic variation in the deamination of gemcitabine and its monophosphate might contribute to variation in therapeutic response to this antineoplastic agent.

Gemcitabine (2′,2′-difluorodeoxycytidine) is a deoxycytidine analogue with antineoplastic activity against solid tumors (1). The mechanism of action of gemcitabine parallels that of 1-β-D-arabinofuranosylcytosine and is thought to require the transport of gemcitabine into cells, followed by metabolic activation to form a triphosphate that is both incorporated into DNA and competitively inhibits DNA synthesis (1). The reactions involved in metabolic activation are depicted graphically in Fig. 1, which also shows gemcitabine inactivation by deamination catalyzed by cytidine deaminase (CDA) as well as the deamination of gemcitabine monophosphate, which is catalyzed by deoxycytidylate deaminase (DCTD). In the present study, we set out to study gemcitabine pharmacogenomics by applying a "genotype-to-phenotype" strategy to CDA and DCTD. That process began by resequencing the genes encoding these two proteins using DNA samples from subjects of differing ethnicity, followed by the creation of expression constructs for all nonsynonymous coding single nucleotide polymorphisms (cSNP) observed and reporter gene constructs for 5′-flanking region polymorphisms. Functional genomic studies were then conducted by transfecting mammalian cells with those constructs.

We observed 17 *CDA* polymorphisms, including one common nonsynonymous cSNP that was present in DNA from both African American and Caucasian American subjects and resulted in a Lys27Gln change in the encoded amino acid sequence. This polymorphism had previously been seen (2–5) and was reported to result in either no change in activity with 1-β-D-arabinofuranosylcytosine (ara-C) as substrate (5) or up

Authors' Affiliations: Departments of [1]Molecular Pharmacology and Experimental Therapeutics and [2]Biochemistry and Molecular Biology, Mayo Clinic College of Medicine, Rochester, Minnesota

Received 9/9/05; revised 11/11/05; accepted 1/4/06.

Grant support: NIH grants R01 GM28157 (O.E. Salavaggione, Y. Ji, and R.M. Weinshilboum), R01 GM35720 (O.E. Salavaggione, Y. Ji, and R.M. Weinshilboum), and U01 GM61388, The Pharmacogenetics Research Network (O.E. Salavaggione, Y. Ji, B.W. Eckloff, L.L. Pelleymounter, E.D. Wieben, and R.M. Weinshilboum), and a grant from Eli Lilly (J.A. Gilbert, M.M. Ames, E.D. Wieben, and R.M. Weinshilboum). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Requests for reprints: Richard M. Weinshilboum, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic College of Medicine, 200 First Street Southwest, Rochester, MN 55905. Phone: 507-284-2246; Fax: 507-284-9111; E-mail: Weinshilboum.Richard@mayo.edu.

doi:10.1158/1078-0432.CCR-05-1969

DOI:10.1158/1078-0432.CCR-05-1969

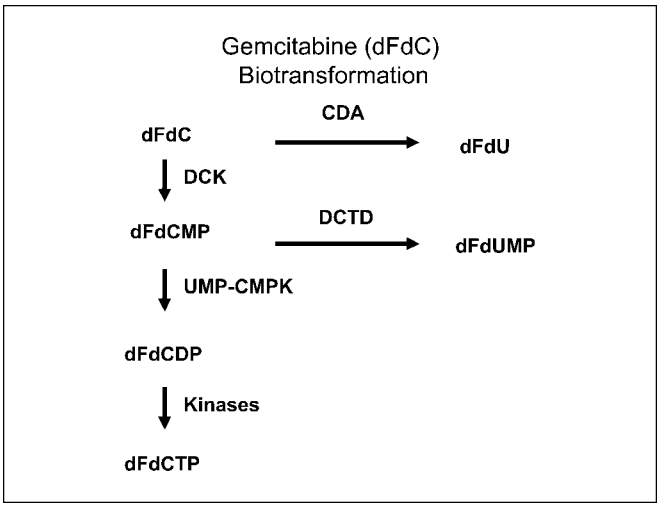

Fig. 1. Gemcitabine biotransformation. DCK, deoxycytidine kinase; UMP-CMPK, deoxycytidylate kinase; dFdC, gemcitabine; dFdCMP, gemcitabine monophosphate; dFdCDP, gemcitabine diphosphate; dFdCTP, gemcitabine triphosphate; dFdU, 2′,2′-difluorodeoxyuridine; and dFdUMP, 2′,2′-difluorodeoxyuridine monophosphate.

to a 3-fold decrease in activity (4). In our hands, transfection of mammalian cells with a Gln27 variant allozyme expression construct resulted in decreased activity relative to wild-type (WT) enzyme with gemcitabine as substrate but without a significant change in the quantity of immunoreactive protein. Furthermore, the variant allozyme had a significantly higher apparent K_m value for gemcitabine than did the WT allozyme. In addition, reporter gene experiments included in the present studies indicated that common polymorphisms present in the 5′-flanking region of *CDA* might influence transcription. Resequencing of *DCTD* showed 29 polymorphisms, including one nonsynonymous cSNP. Functional studies of the Asp58 DCTD variant allozyme showed little basal activity but had elevated levels of immunoreactive protein when compared with WT DCTD. Taken together, these observations suggest that inherited variation in the deamination of gemcitabine and its monophosphate might contribute to individual variation in the therapeutic efficacy and/or toxicity of this increasingly important antineoplastic drug.

Materials and Methods

DNA samples. DNA from 60 African American and 60 Caucasian American subjects (sample sets HD100AA and HD100CAU) was obtained from the Coriell Cell Repository (Camden, NJ). These samples had been deposited by the National Institute of General Medical Sciences and all subjects had provided written consent for the use of their DNA for research purposes. Further information with regard to these anonymized samples can be obtained from the Coriell Cell Repository. The present studies were reviewed and approved by the Mayo Clinic Institutional Review Board.

CDA and DCTD resequencing. All CDA and DCTD exons and splice junctions, as well as a portion of the 5′-flanking regions for each gene, were amplified by use of the PCR. Primers were designed to hybridize within introns, in the 5′-flanking regions or in the 3′-untranslated regions (UTR), of the terminal exon at locations selected to avoid repetitive sequence. All primer sequences are listed in Supplementary data. Amplification reactions were done with AmpliTaq Gold DNA polymerase (Perkin-Elmer, Foster City, CA) with a "hot start." Because of the high GC content of *DCTD* exon 1, PCR reactions for that exon used as template a long PCR product that included ~500 bp of 5′-flanking region, plus exon 1 and intron 1. All PCR amplifications were done in a Perkin-Elmer model 9700 thermal cycler. Amplicons were sequenced in the Mayo Molecular Core Facility on both strands with an ABI 3700 DNA sequencer using BigDye (Perkin-Elmer) dye primer-sequencing chemistry (6). To exclude the possibility of PCR-induced artifacts, independent amplifications, followed by sequencing,

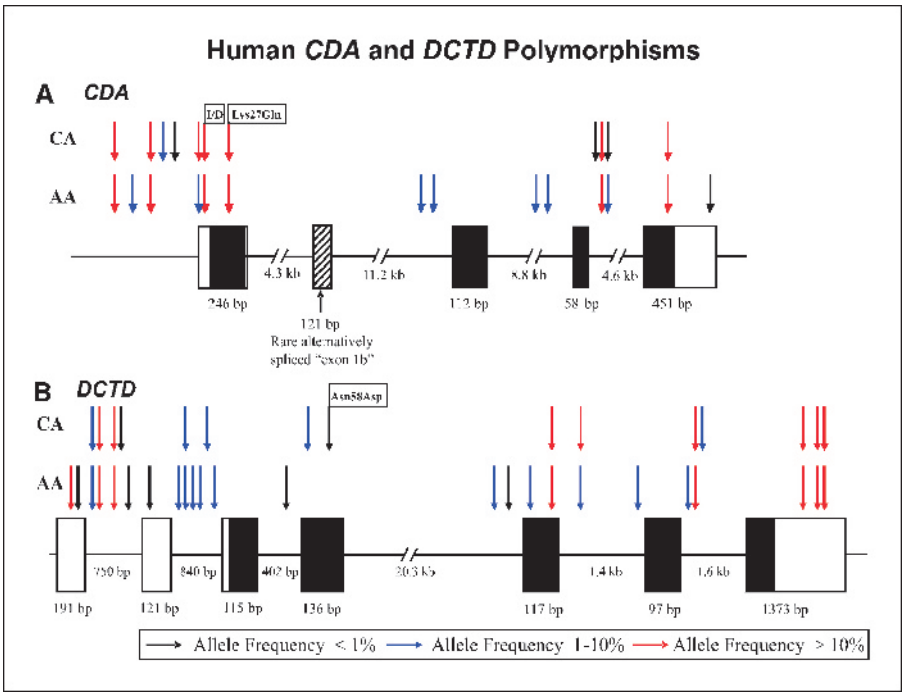

Fig. 2. *CDA* and *DCTD* polymorphisms. *A, CDA* gene structure; arrows, locations of polymorphisms. Black rectangles, open reading frame; white rectangles, portions of exons that encode UTR sequence. Cross-hatched box, a rare, out-of-frame alternatively spliced exon. A change in encoded amino acid resulting from the presence of a nonsynonymous cSNP is also indicated. I/D, an insertion/deletion event. *B,* the *DCTD* gene structure; arrows, locations of polymorphisms.

DOI:10.1158/1078-0432.CCR-05-1969

Table 1. Human *CDA* polymorphisms

Gene location	Nucleotide	Sequence change	Amino acid change	Frequency of variant allele	
				AA	CA
5′-FR	−451	C>T		0.102	0.397
5′-FR	−378	T>C		0.042	0.000
5′-FR	−205	C>G		0.102	0.491
5′-FR	−182	G>A		0.000	0.093
5′-FR	−116	G>T		0.000	0.009
5′-UTR	**−92**	**A>G**		**0.100**	**0.351**
5′-UTR	**−31**	**I/D**		**0.457**	**0.412**
Exon 1	**79**	**A>C**	**Lys27Gln**	**0.108**	**0.298**
IVS 1	−129	G>T		0.025	0.000
IVS 1	−21	T>C		0.017	0.000
IVS 2	−165	A>T		0.050	0.000
IVS 2	−77	A>T		0.083	0.000
IVS 3	64	C>A		0.000	0.008
IVS 3	71	T>C		0.275	0.175
IVS 3	97	G>A		0.017	0.008
Exon 4	**435**	**C>T**		**0.367**	**0.325**
3′-UTR	**543**	**G>A**		**0.008**	**0.000**

NOTE: Gene locations, nucleotide sequence alterations, amino acid sequence alterations, and frequencies of polymorphisms in the two populations studied are listed. Polymorphisms within exons are set in boldface. I/D refers to an insertion-deletion event in which a C was deleted at position (−31) in the most common allele, and IVS is intervening sequence (intron). The numbering scheme for nucleotide positions is based on assignment of the (+1) position to the "A" in the translation initiation codon, with nucleotides 5′ to that position assigned negative and those 3′ within the cDNA assigned positive numbers. Intron nucleotide positions are numbered from the nearest splice site, with +1 as the first nucleotide at the 5′-end of the introns and −1 as the first nucleotide at the 3′-end.
Abbreviation: AA, African American; CA, Caucasian American; FR, flanking region.

were done for any SNP observed in only a single DNA sample or for any sample with an ambiguous chromatogram. The sequencing chromatograms were analyzed using the PolyPhred 3.0 (7), Consed 8.0 (8), and Mutation Surveyor 2.41 (9) programs. The University of Wisconsin GCG software package, version 10, was also used to analyze nucleotide sequence. NT_004610 and NM_001785 were the GenBank accession numbers for the *CDA* reference sequences used in these experiments, and NT_022792 and NM_001921 were the reference sequences for *DCTD*.

CDA and DCTD expression constructs and transient expression. Site-directed mutagenesis done with the QuikChange kit (Stratagene, La Jolla, CA) was used to create variant expression constructs or, in the case of CDA, the WT construct because the Gln27 variant was amplified initially. Sequences of the primers used for site-directed mutagenesis are also listed in Supplementary data. The sequences of inserts were confirmed by sequencing after cloning into the eukaryotic expression vector pCR3.1 (Invitrogen, Carlsbad, CA). Expression constructs for the WT and variant CDA allozyme cDNAs, as well as "empty" pCR3.1 that lacked an insert, were transfected into COS-1 cells in serum-free DMEM (BioWhittaker, Walkersville, MD) using the TransFast reagent (Promega, Madison, WI) at a charge ratio of 1:1. The same approach was used with the DCTD constructs except for the fact that those constructs were transfected into Chinese hamster ovary K1 cells because these cells had negligible DCTD and deoxycytidylate kinase background levels of activity. Specifically, 7 μg of construct DNA were cotransfected with 7 μg of pSV-β-galactosidase DNA (Promega) as a control to make it possible to correct for transfection efficiency. The cells were harvested, homogenized, and cytosol preparations were made as described elsewhere (10).

CDA reporter gene constructs. Luciferase reporter gene constructs were created to study the possible effects of 5′-flanking region polymorphisms on the regulation of *CDA* transcription. Specifically, the PCR was used to amplify DNA segments from samples with known 5′-flanking region haplotypes, and those amplicons were used to create the constructs. Forward and reverse primers included *Mlu*I and *Xho*I restriction sites at their 5′-ends, respectively. Amplification products were digested overnight with these restriction enzymes, followed by purification with the QIAquick PCR Purification Kit (Qiagen, Valencia, CA). The digested products were then cloned into pGL3-basic (Promega), 5′ of the firefly luciferase open reading frame. Each insert was sequenced, and these constructs were used to transfect HEK293T and PC-3 prostate carcinoma cells, both of which showed CDA protein expression by Western blot analysis (data not shown). Specifically, 1 μg of purified plasmid DNA was transfected into the cultured cells with 200 ng of pRL-TK (Promega) DNA that encoded *Renilla* luciferase. The *Renilla* luciferase activity expressed by pRL-TK was used as a control for transfection efficiency. The cells were also transfected with pGL3-basic that lacked insert. Transfections were done using the FuGENE reagent (Roche, Indianapolis, IN). The cells were incubated for 48 hours, followed by harvest in passive lysis buffer (Promega). These cell lysates were used to assay luciferase activity with a Dual-Luciferase Reporter Assay (Promega), and results were expressed as the ratio of firefly luciferase light units to *Renilla* luciferase light units.

Enzyme assays and substrate kinetics. CDA catalyzes the formation of 2′,2′-difluorodeoxyuridine from gemcitabine (Fig. 1). CDA activity was measured with gemcitabine (Eli Lilly, Indianapolis, IN) as substrate using a modification of the assay of Miwa et al. (11) with the high-performance liquid chromatography analysis of Freeman et al. (12). Basal levels of enzyme activity for WT and variant allozymes were corrected for transfection efficiency on the basis of β-galactosidase activity measured with the β-Galactosidase Enzyme Assay System (Promega).

DCTD catalyzes the formation of 2′,2′-difluorodeoxyuridine monophosphate from gemcitabine monophosphate (Fig. 1). A modification of the assay described by Heinemann et al. (13) was used to measure

DOI:10.1158/1078-0432.CCR-05-1969

DCTD activity with the extraction procedure and high-performance liquid chromatography separation described by van Haperen et al. (14). This assay measured the "disappearance" of gemcitabine monophosphate (Eli Lilly) and was optimized with recombinant human WT DCTD that had been expressed in COS-1 cells.

Western blot analysis. Levels of immunoreactive recombinant protein, corrected for transfection efficiency, were analyzed by quantitative Western blot analysis. A rabbit polyclonal antibody to the COOH-terminal 14 amino acids of CDA (15) was provided by Dr. Richard L. Momparler (Universite de Montreal, Hopital Sainte-Justine, Montreal, Quebec, Canada). Rabbit polyclonal antibody to purified DCTD was provided by Dr. Frank Maley (Wadsworth Center, New York State Department of Health, Albany, NY; refs. 16, 17). Specifically, samples for Western blot analysis were loaded on a 15% acrylamide gel (Criterion, Bio-Rad, Hercules, CA) on the basis of the cotransfected β-galactosidase activity to correct for possible variation in transfection efficiency. Immunoreactive proteins were detected with the Enhanced Chemiluminescence System (Amersham) and Hyperfilm-Enhanced Chemiluminescence System (Amersham), and the results were quantitated using the NIH Image program (http://rsb.info.nih.gov/nih-image/).

Data analysis. Linkage analysis was done by testing all possible pairwise combinations of polymorphisms and calculating D' values, a method for reporting linkage data that is independent of allele frequency (18, 19). These data were then depicted graphically. Haplotypes were inferred computationally by using a program based on the Expectation-Maximization algorithm (20–22). Apparent K_m values were calculated by nonlinear least squares regression analysis done with the GraphPad Prism program (GraphPad Software, San Diego, CA).

Results

CDA and DCTD gene resequencing. Before resequencing, 10 kb of human *CDA* genomic DNA sequence located 5′ of the ATG translation initiation codon were compared with the mouse genome sequence. The CDA cDNA open reading frame was also used to search the human EST database. Comparison with the mouse genome resulted in the identification of an area 377 bp in length in the human genome that was located 90 bp 5′ of *CDA* exon 1 that was 70.6% identical to the mouse genome sequence, compatible with the conclusion that this 5′-flanking region sequence probably included the core promoter for the gene. Our search of the human EST database

Table 2. Human *DCTD* polymorphisms

Gene location	Nucleotide	Sequence change	Amino acid change	Frequency of variant allele	
				AA	CA
Exon 1	**−274**	**G>A**		**0.115**	**0.000**
Exon 1	**−256**	**C>G**		**0.009**	**0.000**
IVS 1	148	G>A		0.018	0.042
IVS 1	338	G>A		0.208	0.217
IVS 1	−333	T>C		0.342	0.333
IVS 1	−218	G>A		0.000	0.008
IVS 1	−184	G>A		0.008	0.000
Exon 2	**−126**	**A>G**		**0.009**	**0.000**
IVS 2	9	C>A		0.051	0.000
IVS 2	−517	G>A		0.050	0.025
IVS 2	−353	C>T		0.050	0.000
IVS 2	−169	A>T		0.050	0.000
IVS 2	−106	G>A		0.000	0.033
IVS 2	−35	C>T		0.025	0.000
IVS 3	−31	G>A		0.008	0.000
Exon 4	**114**	**C>T**		**0.000**	**0.017**
Exon 4	**172**	**A>G**	**Asn58Asp**	**0.000**	**0.008**
IVS 4	−150	G>A		0.017	0.000
IVS 4	−44	G>T		0.008	0.000
Exon 5	**255**	**G>C**		**0.017**	**0.000**
Exon 5	**315**	**T>C**		**0.475**	**0.333**
IVS 5	81	A>G		0.100	0.308
IVS 5	−36	G>A		0.025	0.000
IVS 6	6	T>C		0.058	0.000
IVS 6	15	T>C		0.117	0.300
IVS 6	18	G>A		0.000	0.075
3′-UTR	**738**	**G>A**		**0.242**	**0.617**
3′-UTR	**788**	**A>G**		**0.525**	**0.200**
3′-UTR	**802**	**T>C**		**0.525**	**0.200**

NOTE: Gene locations, nucleotide sequence alterations, amino acid sequence alterations, and frequencies of polymorphisms in the two populations studied are listed. Polymorphisms within exons are set in boldface. The numbering scheme is that described in the legend for Table 1.

DOI:10.1158/1078-0432.CCR-05-1969

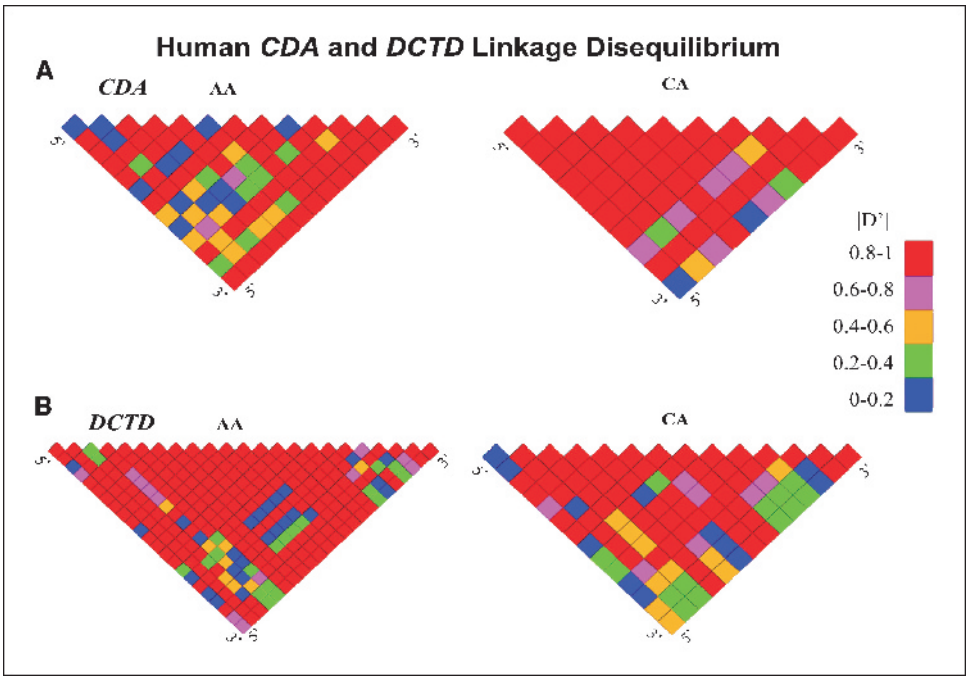

Fig. 3. Human *CDA* and *DCTD* linkage disequilibrium. The extent of population-specific linkage disequilibrium, depicted as pairwise $|D'|$ values, is displayed graphically for both genes.

Table 3. Human *CDA* haplotype analysis

Allele designation	Frequency		5′-FR (−451)	5′-FR (−378)	5′-FR (−205)	5′-FR (−182)	5′-UTR (−92)	5′-UTR (−31)	Exon 1 (79)	IVS 1 (−129)	IVS 1 (−21)	IVS 2 (−165)	IVS 2 (−77)	IVS 3 (71)	Exon 4 (435)
	AA	CA													
*1A	0.372*	0.370*	C	T	C	G	A	***I***	A	G	T	A	A	T	C
*1B	0.186*	0.026	C	T	C	G	A	D	A	G	T	A	A	T	C
*1C	0.091*	0.050	C	T	C	G	A	D	A	G	T	A	A	***C***	***T***
*1D	0.053	0.024*	C	T	C	G	A	***I***	A	G	T	A	A	***C***	***T***
*1E	0.028		C	T	C	G	A	D	A	G	T	A	***T***	T	C
*1F	0.027		C	T	C	G	A	D	A	G	T	***T***	A	***C***	***T***
*1G	0.025*		C	T	C	G	A	D	A	***T***	T	A	A	***C***	***T***
*1H	0.018		C	T	C	G	A	D	A	G	T	A	A	T	***T***
*1I	0.015		C	***C***	C	G	A	D	A	G	T	A	***T***	***C***	***T***
*1J	0.011*		C	T	C	G	A	***I***	A	G	***C***	A	A	T	C
*1K		0.062	C	T	***G***	***A***	A	D	A	G	T	A	A	***C***	***T***
*1L		0.026	***T***	T	***G***	G	***G***	D	A	G	T	A	A	T	C
*1M		0.023	***T***	T	***G***	G	***G***	D	A	G	T	A	A	T	***T***
*1N		0.022	***T***	T	***G***	G	A	***I***	A	G	T	A	A	T	C
*1O		0.022	***T***	T	***G***	G	A	D	A	G	T	A	A	T	C
*1P		0.021	C	T	***G***	***A***	A	D	A	G	T	A	A	T	C
*1Q		0.012	C	T	C	G	***G***	D	A	G	T	A	A	T	C
*2A	0.041	0.116*	***T***	T	***G***	G	***G***	D	***C***	G	T	A	A	T	***T***
*2B	0.013	0.167*	***T***	T	***G***	G	***G***	D	***C***	G	T	A	A	T	C
*2C		0.014	***T***	T	***G***	G	***G***	D	***C***	G	T	A	A	***C***	***T***
*2D		0.010	C	T	C	G	A	***I***	***C***	G	T	A	A	T	***T***

NOTE: Inferred haplotypes with frequencies ≥1% are also listed. Variant nucleotide sequence is set in boldface and italics. The WT allozyme has been designated *1 and letter designations were then added based on descending allele frequencies, starting with haplotypes present in African American subjects. The *2 designation was used for sequences that encoded Gln27. The symbols I and D at 5′-UTR location (−31) represent insertion and deletion, respectively. Polymorphisms at positions (−116), I3 (64), I3 (97), and 543 were excluded from this table because they were not represented in any of the haplotypes listed.

*Frequencies of the eight unambiguous haplotypes observed.

DOI:10.1158/1078-0432.CCR-05-1969

Table 4. Human *DCTD* haplotype analysis

Allele designation	Frequency		Exon 1 (−274)	IVS 1 (148)	IVS 1 (338)	IVS 1 (−333)	IVS 2 (9)	IVS 2 (−517)	IVS 2 (−353)	IVS 2 (−169)
	AA	CA								
**1A*	0.214*	0.478*	G	G	G	T	C	G	C	A
**1B*	0.116*	0.045	G	G	G	T	C	G	C	A
**1C*	0.107*		G	G	G	T	C	G	C	A
**1D*	0.083*	0.009*	G	G	G	T	C	G	C	A
**1E*	0.050		***A***	G	G	***C***	***A***	***A***	***T***	***T***
**1F*	0.05*		G	G	***A***	***C***	C	G	C	A
**1G*	0.034		G	G	G	T	C	G	C	A
**1H*	0.026		G	G	G	***C***	C	G	C	A
**1I*	0.025	0.027	G	G	***A***	***C***	C	G	C	A
**1J*	0.024*		G	G	G	***C***	C	G	C	A
**1K*	0.022		G	G	***A***	***C***	C	G	C	A
**1L*	0.019		G	G	***A***	***C***	C	G	C	A
**1M*	0.013		***A***	G	***A***	***C***	C	G	C	A
**1N*	0.011		***A***	G	***A***	***C***	C	G	C	A
**1O*	0.011		G	G	G	T	C	G	C	A
**1P*		0.099*	G	G	***A***	***C***	C	G	C	A
**1Q*		0.060	G	G	***A***	***C***	C	G	C	A
**1R*		0.042	G	G	G	***C***	C	G	C	A
**1S*		0.039	G	G	G	T	C	G	C	A
**1T*		0.028	G	G	G	T	C	G	C	A
**1U*		0.025	G	G	G	***C***	C	G	C	A
**1V*		0.017	G	***A***	G	T	C	***A***	C	A
**1W*		0.015*	G	G	G	***C***	C	G	C	A
**1X*		0.011	G	G	G	T	C	G	C	A
**1Y*		0.010*	G	G	G	***C***	C	G	C	A
**2A*		0.008	G	G	G	T	C	G	C	A

NOTE: Inferred haplotypes with frequencies ≥1% are also listed. Variant nucleotide sequence is set in boldface and italics. The WT allozyme has been designated as **1* and letter designations were then added based on descending allele frequencies, starting with haplotypes present in African American subjects. Although we were unable to determine the haplotype unequivocally, the **2* designation was used to refer to sequence that encoded Asp58. Polymorphisms at positions −256, I1 (−218), I1 (−184), −126, I2 (−35), I3 (−31), I4 (−150), I4 (−44), 255, I5 (−36), and I6 (6) were excluded from this table because they were not represented in any of the haplotypes listed.
*Frequencies of the nine unambiguous haplotypes observed.

indicated that CDA mRNA was expressed in human bone marrow, prostate, pancreas, liver, and spleen. In addition, a small number of EST sequences, obtained mainly from leukocytes, included an out-of-frame, rarely represented, alternatively spliced exon (referred to here as "exon 1b") located between coding exons 1 and 2 (see Fig. 2A).

CDA was resequenced in 120 DNA samples. When this sequence was compared with the consensus *CDA* sequence, 17 polymorphisms, including 16 SNPs and 1 insertion-deletion, were observed in the 240 alleles resequenced (Fig. 2A; Table 1). Fourteen polymorphisms were present in DNA samples from African American and 11 in DNA from Caucasian American subjects. Six of the SNPs in DNA from the African American subjects were not observed in the Caucasian American subjects, and three of those present in the Caucasian American samples were absent from the African American DNA samples. Thirteen of the 14 polymorphisms in DNA from African American subjects and 8 of the 11 in DNA from Caucasian American subjects had frequencies of >1%, and, as a result, would be considered "common" in these populations. One nonsynonymous cSNP, 79 A>C (Lys27Gln), was observed, with allele frequencies of 10.8% and 29.8% in African American and Caucasian American subjects, respectively (Fig. 2A; Table 1). The 208 G>A (Ala70Thr) polymorphism previously reported in Japanese and African American subjects (5, 23) was not observed in any of our samples. We also compared the polymorphisms that we observed with those deposited in publicly accessible databases. A total of 11 of the 17 polymorphisms that were identified in this study had been deposited in dbSNP build 124 (http://www.ncbi.nlm.nih.gov/SNP), including the nonsynonymous cSNP that resulted in the Lys27Gln change in encoded amino acid. Four of the six polymorphisms not represented in dbSNP were observed only in the African American DNA samples.

A search of the human EST database before resequencing *DCTD* resulted in the identification of numerous transcripts that included sequence located ~1.0 and 1.9 kb 5′-upstream from the ATG start codon. Furthermore, these noncoding exons were bounded by appropriately placed "GT-AG" canonical splice sequences (24). As a result, we also resequenced these upstream noncoding exons, as well as ~500 bp of sequence located immediately upstream of the initial noncoding exon,

DOI:10.1158/1078-0432.CCR-05-1969

Table 4. Human *DCTD* haplotype analysis (Cont'd)

IVS 2 (−106)	Exon 4 (114)	Exon 4 (172)	Exon 5 (315)	IVS 5 (81)	IVS 6 (15)	IVS 6 (18)	3′-UTR (738)	3′-UTR (788)	3′-UTR (802)
G	C	A	T	A	T	G	***A***	***A***	***T***
G	C	A	T	A	T	G	G	G	C
G	C	A	***C***	A	T	G	G	G	C
G	C	A	T	A	T	G	G	***A***	***T***
G	C	A	T	A	T	G	G	G	C
G	C	A	***C***	A	T	G	G	G	C
G	C	A	***C***	A	T	G	G	***A***	***T***
G	C	A	***C***	A	T	G	G	G	C
G	C	A	***C***	***G***	***C***	G	G	G	C
G	C	A	***C***	A	T	G	G	***A***	***T***
G	C	A	***C***	***G***	T	G	G	***A***	***T***
G	C	A	T	A	T	G	G	G	C
G	C	A	***C***	A	T	G	G	G	C
G	C	A	***C***	***G***	T	G	G	***A***	***T***
G	C	A	***C***	A	T	G	***A***	***A***	***T***
G	C	A	***C***	***G***	***C***	G	G	***A***	***T***
G	C	A	***C***	***G***	***C***	G	***A***	***A***	***T***
G	C	A	T	A	T	***A***	G	***A***	***T***
G	C	A	***C***	***G***	***C***	G	G	G	C
G	C	A	***C***	***G***	***C***	G	G	***A***	***T***
G	C	A	T	A	T	***A***	G	G	C
A	C	A	***C***	A	T	G	G	G	C
G	C	A	T	A	T	G	G	G	C
G	***T***	A	T	A	T	G	***A***	***A***	***T***
G	C	A	T	A	T	G	***A***	***A***	***T***
G	C	***G***	***C***	***G***	***C***	G	***A***	G	C

plus the five *DCTD* coding exons and all splice junctions. *DCTD* was also resequenced in 120 DNA samples. When those data were compared with consensus *DCTD* sequence, 29 polymorphisms were observed (Fig. 2B; Table 2). Twenty-four of these polymorphisms were present in DNA samples from African American, with 15 in DNA from Caucasian American subjects. Fourteen of the SNPs in the African American subjects were not observed in the Caucasian American subjects, and five of those present in the Caucasian American samples were absent from African American subjects. Nineteen of the 24 polymorphisms in DNA from African American subjects and 13 of the 15 in DNA from Caucasian American subjects had frequencies >1% and, as a result, would be considered common in these populations. One nonsynonymous cSNP, 172 A>G (Asn58Asp), was observed in a single Caucasian American subject (Fig. 2B; Table 2). No SNPs were observed in the 500 bp immediately upstream of *DCTD*. We also compared the *DCTD* SNPs with those in dbSNP build 124. A total of 11 of the 29 polymorphisms that we observed had been deposited in dbSNP. Of the 18 polymorphisms that were not in the database, 12 were found only in the African American and 4 were observed only in Caucasian American DNA samples.

Linkage disequilibrium and haplotype analysis. Linkage disequilibrium analysis was done for all of the polymorphisms that we observed in *CDA* and *DCTD*. Specifically, *D′* values were calculated for pairwise combinations of all polymorphisms in each gene (18, 19). The results for both African American and Caucasian American subjects are depicted graphically in Fig. 3. Because haplotype is very useful for molecular epidemiology and pharmacogenomic studies (25), we also did haplotype analyses for both genes (Tables 3 and 4). Eight unambiguous *CDA* haplotypes (i.e., those deduced entirely on the basis of genotype) were observed in the 120 DNA samples resequenced, including four that were unique to African American and three to Caucasian American subjects (Table 3). These unequivocal haplotypes accounted for 69% and 68% of all samples in the two ethnic groups, respectively. The Expectation-Maximization algorithm (20–22) was used to infer a total of 25 and 16 additional *CDA* haplotypes for DNA samples from African American and Caucasian American subjects, respectively. However, only inferred haplotypes with frequencies ≥1% have been listed in Table 3. Information with regard to inferred haplotypes with frequencies <1% can be obtained from the authors on request. Haplotype designations were made initially on the basis of the encoded amino acid sequence, with the WT amino acid sequence being designated **1*. We then assigned letter designations based on descending allele frequencies, starting with haplotypes present in African American subjects. Specifically, **1A* was more common than **1B*, etc. (Table 3). **2* haplotypes for *CDA* encoded the common Gln27 variant allozyme. A similar analysis was done for *DCTD*, and nine unambiguous haplotypes were observed in the 120 DNA samples studied, including six in African American and five in Caucasian American subjects (Table 4).

DOI:10.1158/1078-0432.CCR-05-1969

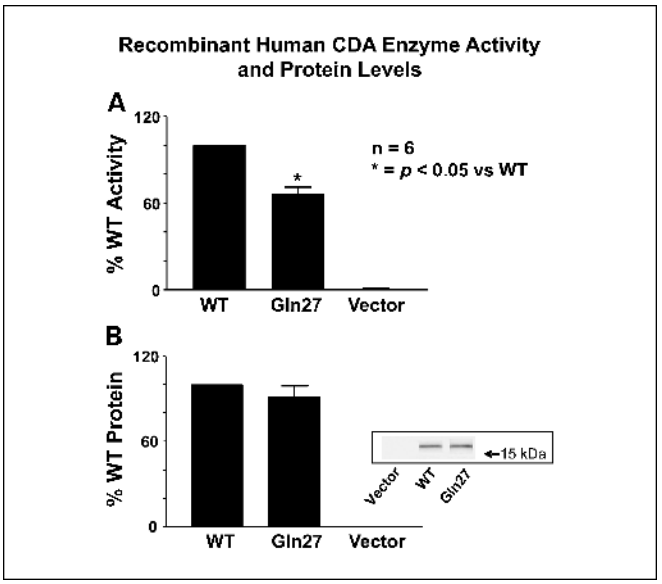

Fig. 4. Recombinant human CDA allozyme activity and immunoreactive protein. *A,* columns, mean enzyme activity for recombinant allozymes and an empty vector control, corrected for transfection efficiency ($n = 6$); bars, SE. *, $P < 0.05$, compared with WT. *B,* mean immunoreactive protein corrected for transfection efficiency ($n = 6$). Inset, representative Western blot used to obtain the data shown in (*B*).

Unequivocal *DCTD* haplotypes accounted for 59% and 61% of all samples from the two ethnic groups, respectively. There was also a total of 36 and 22 inferred haplotypes for African American and Caucasian American subjects, respectively, and those with frequencies of ≥1% are listed in Table 4. *DCTD* haplotype designations were made using a process similar to that used for *CDA*, with **2* indicating the haplotype that encoded the Asp58 variant allozyme.

Functional genomic studies. To characterize possible functional consequences of alterations in encoded amino acid sequence resulting from the nonsynonymous SNPs in *CDA* and *DCTD*, mammalian expression constructs were created for the WT and variant allozymes for both genes. Those constructs were used to transfect mammalian cells, and basal levels of enzyme activity in cytosolic preparations were measured under optimal conditions. In each case, two rounds of transfection were done, each in triplicate, and levels of enzyme activity or protein were corrected for the cotransfected β-galactosidase to correct for possible variation in transfection efficiency. Average levels of basal enzyme activity for the two CDA allozymes studied are depicted graphically in Fig. 4A. The Gln27 variant had 66 ± 5.1% (mean ± SE) of the WT activity ($P < 0.05$). When substrate kinetic studies were done, the apparent K_m value of the Gln27 allozyme, 397 ± 40 μmol/L, was significantly higher than that for WT CDA, 289 ± 20 μmol/L; $P < 0.025$ (Fig. 5). When recombinant allozyme protein concentrations, corrected for transfection efficiency, were measured by quantitative Western blot analysis, immunoreactive protein levels did not differ significantly between WT and Gln27 CDA allozymes (Fig. 4B). Expression constructs were also created for the DCTD WT and Asp58 variant allozymes. Recombinant Asp58 DCTD had very little enzyme activity but levels of immunoreactive protein for this allozyme were significantly elevated when compared with those for the WT enzyme (Fig. 6). Unfortunately, because of the very low level of activity for the Asp58 variant allozyme, substrate kinetic studies were not possible.

The final series of functional genomic studies involved an attempt to determine whether polymorphisms located in the 5′-flanking region of *CDA* might influence transcription. Because we had not detected polymorphisms in the 5′-flanking region of *DCTD*, reporter gene studies were limited to *CDA*. Specifically, reporter gene constructs were created for common *CDA* 5′-flanking region haplotypes (Table 5). A total of 10 different constructs were transfected into HEK293T and PC-3 cells. Reporter gene activity showed similar patterns in both cell lines (Fig. 7). There were significant differences among 5′-flanking region haplotypes in their ability to drive transcription.

Discussion

Gemcitabine can be inactivated by deamination catalyzed by CDA and DCTD. It also undergoes phosphorylation-dependent metabolic activation, which converts the drug to an active triphosphate nucleotide (Fig. 1). In the present study, we set out to determine the nature and extent of sequence variation in *CDA* and *DCTD* and to do initial functional genomic experiments as a step toward pharmacogenetic studies of this important antineoplastic drug. We began by resequencing exons, splice junctions, and a portion of the 5′-flanking region of these two genes using DNA samples from both African American and Caucasian American subjects. For *CDA*, 17

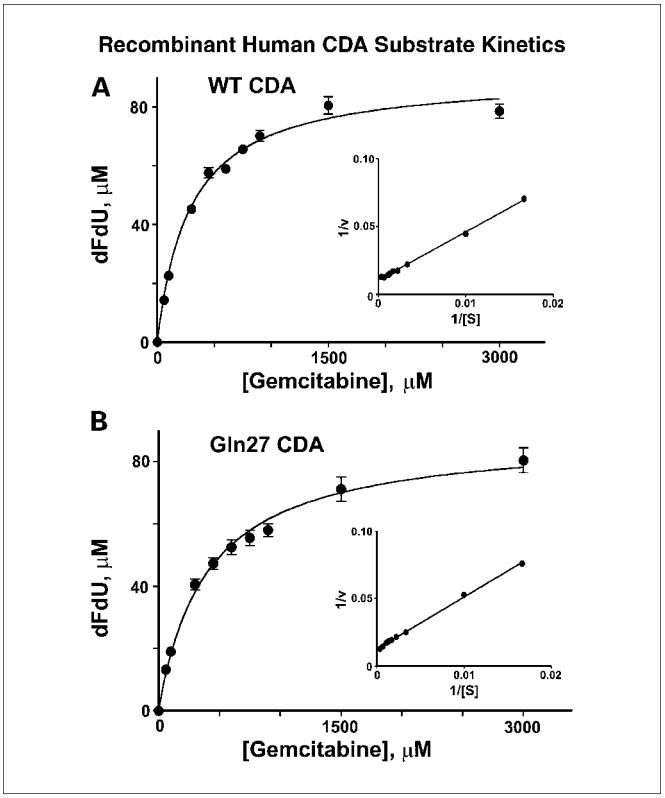

Fig. 5. Recombinant CDA substrate kinetics. Gemcitabine substrate kinetics for the WT (*A*) and Gln27 (*B*) allozymes ($n = 6$). Points, mean; bars, SE.

DOI:10.1158/1078-0432.CCR-05-1969

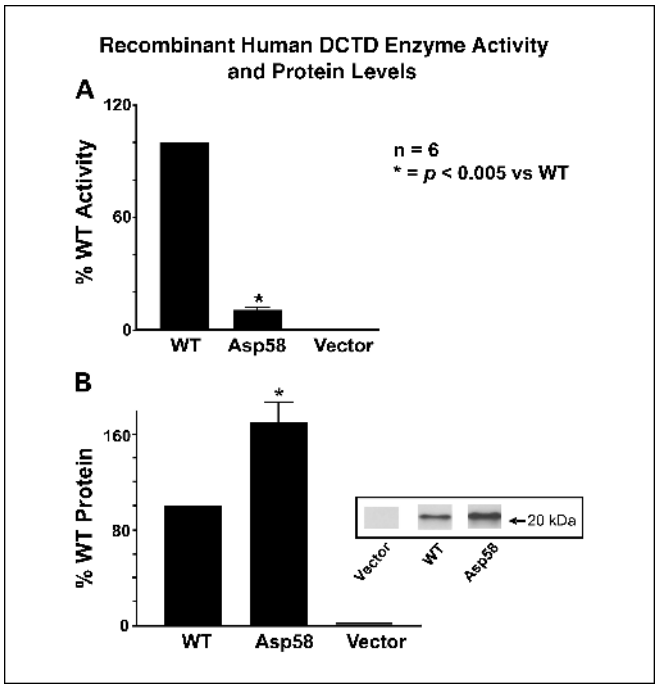

Fig. 6. Recombinant human DCTD allozyme activity and immunoreactive protein. *A,* columns, mean enzyme activity for recombinant allozymes and an empty vector control, corrected for transfection efficiency ($n = 6$); bars, SE. *, $P < 0.005$, compared with WT. *B,* mean immunoreactive protein corrected for transfection efficiency ($n = 6$). *, $P < 0.005$, compared with WT. Inset, representative Western blot used to obtain the data shown in (*B*).

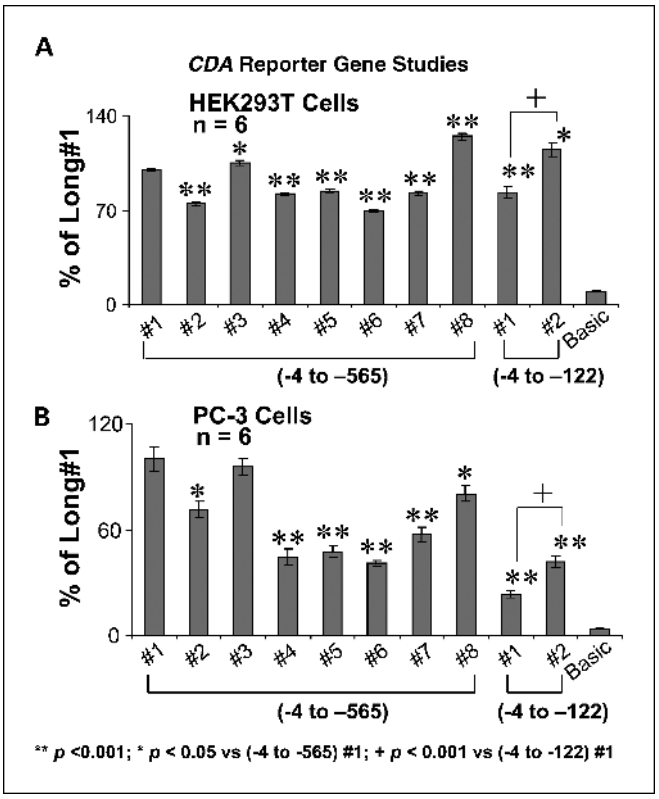

Fig. 7. *CDA* 5′-flanking region luciferase reporter gene studies. Luciferase activity levels for reporter gene constructs containing *CDA* 5′-flanking region haplotypes transfected into HEK293T (*A*) or PC-3 (*B*) cells. Construct names refer to the 5′-flanking region haplotype designations listed in Table 5. 5′-Flanking region insert lengths are also indicated. Columns, mean ($n = 6$); bars, SE.

polymorphisms, including 1 nonsynonymous cSNP, were observed in the 120 DNA samples studied (Table 1; Fig. 2A) whereas 29 *DCTD* polymorphisms, including 1 novel nonsynonymous cSNP, were observed (Table 2; Fig. 2B). For both genes, more SNPs were present in DNA samples from African American than in those from Caucasian American subjects. This observation is compatible with previous reports of greater DNA sequence diversity in African populations than in other ethnic groups (26). Finally, a series of haplotypes were observed for both genes in each population (Tables 3 and 4).

The next step in our studies involved functional genomic experiments. Expression constructs were created for the nonsynonymous cSNPs that we observed in both *CDA* and *DCTD*, and those constructs were used to transiently transfect mammalian cells. The common *CDA* 79 A>C (Lys27Gln) polymorphism resulted in a moderate decrease in the level of CDA activity that was associated with a modest alteration in apparent K_m value. The functional consequences of this

Table 5. Human *CDA* 5′-FR haplotypes used in the reporter gene experiments

Reporter gene constructs		Frequency		(−451)	(−378)	(−205)	(−182)	(−116)	(−92)	(−31)
(−4 to −565)	(−4 to −122)	AA	CA							
1	1	0.455	0.394	C	T	C	G	G	A	***I***
2	2	0.386	0.096	C	T	C	G	G	A	D
3	—	0.083	0.346	***T***	T	***G***	G	G	***G***	D
4	—	0.042	—	C	***C***	C	G	G	A	D
5	—	0.017	0.009	C	T	C	G	G	***G***	D
6	—	0.017	0.026	***T***	T	***G***	G	G	A	D
7	—	—	0.092	C	T	***G***	***A***	G	A	D
8	—	—	0.028	***T***	T	***G***	G	G	A	***I***

NOTE: These haplotypes for only the 5′-FR of the gene represent a "collapsed" subset of the data shown in Table 3. Nucleotides in boldface and italics differ from the WT sequence.
Abbreviations: I, insertion of a C; D, deletion of a C at that location.

DOI:10.1158/1078-0432.CCR-05-1969

polymorphism had been studied previously (4, 5). Kirch et al. (4) reported that deamination of ara-C by bacterial recombinant WT CDA was 1.3- to 3.3-fold that of Gln27. Yue et al. (5) reported an allele frequency of 20.1% for the C79 allele in a study of 221 Japanese subjects but did not observe a significant difference in enzyme activity between the variant and WT allozymes after expression in yeast with either cytidine or ara-C as substrates. Our apparent K_m value of 289 μmol/L for recombinant WT CDA with gemcitabine as a substrate can be compared with values of 95.7 μmol/L for purified human placental CDA (27) and 264 μmol/L for bacterially expressed recombinant human CDA (11), both measured with the same substrate that we used. In our studies, the DCTD Asp58 allozyme exhibited very little enzyme activity but did have an elevated level of immunoreactive protein when compared with the WT enzyme (Fig. 6). These observations raise the possibility that this alteration in amino acid sequence might disrupt the ability of the protein to catalyze the enzyme reaction. However, because of the very low level of activity for this variant allozyme, substrate kinetic studies were not possible. Although this variant is rare, its striking effect on DCTD activity indicates that it might be associated with a clinically relevant phenotype, so its clinical implications should be pursued in future clinical pharmacogenetic studies. The mechanism(s) responsible for the alterations in levels of immunoreactive protein for CDA Gln27 and DCTD Asp58 should be explored in the course of future studies. We also did reporter gene experiments to study the possible effect of *CDA* 5′-flanking region variant haplotypes on transcription. Patterns of reporter gene expression were similar in the two cell lines studied but there were significant differences among 5′-flanking region haplotypes in their ability to drive transcription (Fig. 7). Demontis et al. (28) have reported that the *CDA* 5′-flanking region contains numerous potential transcription factor binding sites, so the present experiments represent merely an initial step in defining mechanisms responsible for the variations in transcription that we observed.

In summary, we have characterized the nature and extent of common sequence variation in the human *CDA* and *DCTD* genes in two ethnic groups. We also did functional studies of this sequence variation. The *in vivo* functional and clinical implications of these observations remain to be determined, as do the possible functional consequences of the other polymorphisms that we observed. However, the present results raise the possibility that ethnic-specific genetic variation in *CDA* and *DCTD* might contribute to individual differences in therapeutic response to treatment with gemcitabine.

Acknowledgments

We thank Luanne Wussow for her assistance with the preparation of this manuscript.

References

1. Chabner BA. Cytidine analogues. In: Chabner BA, Longo DL, editors. Cancer chemotherapy and biotherapy: principles and practice. 2nd ed. Philadelphia and New York: Lippincott-Raven; 1996. p 213–33.
2. Kuhn K, Bertling WM, Emmrich F. Cloning of a functional cDNA for human cytidine deaminase (CDD) and its use as a marker of monocyte/macrophage differentiation. Biochem Biophys Res Commun 1993; 190:1–7.
3. Watanabe S, Uchida T. Expression of cytidine deaminase in human solid tumors and its regulation by 1 α,25-dihydroxyvitamin D3. Biochim Biophys Acta 1996;1312:99–104.
4. Kirch H-C, Schroder J, Hoppe H, Esche H, Seeber S, Schutte J. Recombinant gene products of two natural variants of the human cytidine deaminase gene confer different deamination rates of cytarabine *in vitro*. Exp Hematol 1998;26:421–5.
5. Yue L, Saikawa Y, Ota K, et al. A functional single-nucleotide polymorphism in the human cytidine deaminase gene contributing to ara-C sensitivity. Pharmacogenetics 2003;13:29–38.
6. Chadwick RB, Conrad MP, McGinnis MD, Johnston-Dow L, Spurgeon SL, Kronick MN. Heterozygote and mutation detection by direct automated fluorescent DNA sequencing using a mutant Taq DNA polymerase. BioTechniques 1996;20:676–83.
7. Nickerson DA, Tobe VO, Taylor SL. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 1997;25: 2745–51.
8. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. Genome Res 1998;8: 195–202.
9. Bardelli A, Parsons DW, Silliman N, et al. Mutational analysis of the tyrosine kinome in colorectal cancers. Science 2003;300:949.
10. Ji Y, Salavaggione OE, Wang L, et al. Human phenylethanolamine N-methyltransferase pharmacogenomics: gene resequencing and functional genomics. J Neurochem 2005;95:1766–76.
11. Miwa M, Eda H, Ura M, et al. High susceptibility of human cancer xenografts with higher levels of cytidine deaminase to a 2′-deoxycytidine antimetabolite, 2′-deoxy-2′-methylidenecytidine. Clin Cancer Res 1998;4:493–7.
12. Freeman KB, Anliker S, Hamilton M, et al. Validated assays for the determination of gemcitabine in human plasma and urine using high-performance liquid chromatography with ultraviolet detection. J Chromatogr B 1995;665:171–81.
13. Heinemann V, Xu Y-Z, Chubb S, et al. Cellular elimination of 2′,2′-difluorodeoxycytidine 5′-triphosphate: a mechanism of self-potentiation. Cancer Res 1992; 52:533–9.
14. van Haperen VWTR, Veerman G, Boven E, Noordhuis P, Vermorken JB, Peters GJ. Schedule dependence of sensitivity to 2′,2′-difluorodeoxycytidine (gemcitabine) in relation to accumulation and retention of its triphosphate in solid tumour cell lines and solid tumours. Biochem Pharmacol 1994;48:1327–39.
15. Momparler RL, Laliberte J, Eliopoulos N, Beausejour C, Cournoyer D. Transfection of murine fibroblast cells with human cytidine deaminase cDNA confers resistance to cytosine arabinoside. Anticancer Drugs 1996;7:266–74.
16. Weiner KXB, Weiner RS, Maley F, Maley GF. Primary structure of human deoxycytidylate deaminase and overexpression of its functional protein in Escherichia coli. J Biol Chem 1993;17:12983–9.
17. Maley GF, Lobo AP, Maley F. Properties of an affinity-column-purified human deoxycytidylate deaminase. Biochim Biophys Acta 1993;1162: 161–70.
18. Hartl DL, Clark AG. Principles of population genetics. 3rd ed. Sunderland (MA): Sinauer Associates, Inc.; 1997. p. 96–106.
19. Hedrick PW. Genetics of populations. 2nd ed. Sudbury (MA): Jones and Bartlett; 2000. p. 396–405.
20. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 1995;12:921–7.
21. Long JC, Williams RC, Urbanek M. An E-M algorithm and testing strategy for multiple-locus haplotypes. Am J Hum Genet 1995;56:799–810.
22. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 2002;70:425–34.
23. Fukunaga AK, Marsh S, Murry DJ, Hurley TD, McLeod HL. Identification and analysis of single-nucleotide polymorphisms in the gemcitabine pharmacologic pathway. Pharmacogenomics J 2004;4: 307–14.
24. Mount SM. A catalogue of splice junction sequences. Nucleic Acids Res 1982;10:459–72.
25. Drysdale CM, McGraw DW, Stack CB, et al. Complex promoter and coding region β2-adrenergic receptor haplotypes alter receptor expression and predict *in vivo* responsiveness. Proc Natl Acad Sci U S A 2000;97:10483–8.
26. Tishkoff SA, Williams SM. Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet 2002;3:611–21.
27. Bouffard DY, Laliberte J, Momparler RL. Kinetic studies on 2′,2′-difluorodeoxycytidine (gemcitabine) with purified human deoxycytidine kinase and cytidine deaminase. Biochem Pharmacol 1993;45:1857–61.
28. Demontis S, Terao M, Brivio M, Zanotta S, Bruschi M, Garattini E. Isolation and characterization of the gene coding for human cytidine deaminase. Biochim Biophys Acta 1998;1443:323–33.