Functional and Genetic Diversity in the Concentrative Nucleoside Transporter, CNT1, in Human Populations


Received September 2, 2003; accepted November 17, 2003

ABSTRACT

The concentrative nucleoside transporter, CNT1 (SLC28A1), mediates the cellular uptake of naturally occurring pyrimidine nucleosides and many structurally diverse anticancer and antiviral nucleoside analogs. As a first step toward understanding whether genetic variation in CNT1 contributes to variation in the uptake and disposition of clinically used nucleoside analogs, we determined the haplotype structure and functionally analyzed all coding region variants of CNT1 identified in ethnically diverse populations (100 African Americans, 100 European Americans, 30 Asians, 10 Mexican Americans, and 7 Pacific Islanders) (Leabman et al., 2003). A total of 58 coding region haplotypes were identified using PHASE analysis, 44 of which contained at least one amino acid variant. More than half of the coding region haplotypes were population-specific. Using site-directed mutagenesis, 15 protein-altering CNT1 variants, including one amino acid insertion and one base pair (bp) deletion, were constructed and expressed in Xenopus laevis oocytes. All variant transporters took up [3H]thymidine with the exception of CNT1-Ser546Pro, a rare variant, and CNT1–1153del, a single bp deletion found at a frequency of 3% in the African American population. The bp deletion results in a frameshift followed by a stop-codon. The anticancer nucleoside analog gemcitabine had a reduced affinity for CNT1-Val189Ile (a common CNT1 variant found at a frequency of 26%) compared with reference CNT1 (IC50/H11005 13.8/H9262 0.60 M for CNT1-reference and 23.3/H11006 1.5/H9262 M for CNT1-Val189Ile, p < 0.05). These data suggest that common genetic variants of CNT1 may contribute to variation in systemic and intracellular levels of anticancer nucleoside analogs.

Nucleoside analogs are used clinically as anticancer drugs (e.g., cladribine, fludarabine, and gemcitabine) and antiviral drugs (e.g., cytarabine and zalcitabine). These synthetic analogs are modified derivatives of naturally occurring nucleosides and exert their effects by entering nucleoside salvage pathways. Once in salvage pathways, the phosphorylated analogs substitute for naturally occurring nucleotides and inhibit a variety of chemical reactions required for viral or host-cell nucleic acid replication.

Interindividual differences in response to anticancer and antiviral nucleoside analogs represents a major obstacle in drug therapy. Such differences have been associated with variation in systemic and intracellular levels of phosphorylated nucleotide analogs (Abbruzzese, 2002). For example, recent studies have shown that variation in intracellular levels of phosphorylated gemcitabine nucleotides is dependent on the rate of delivery of gemcitabine to cells (Tempero et al., 2003). An optimal rate of entry is critical in maximizing the cytotoxic response to gemcitabine.

To enter cells, naturally occurring nucleosides as well as many synthetic nucleoside analogs require plasma membrane nucleoside transporters. The concentrative nucleoside transporter CNT1 (SLC28A1) is one of three members of the SLC28 family of sodium-dependent nucleoside transporters (Ritzel et al., 1997). In contrast to CNT2, which prefers purine nucleosides, and CNT3, which does not discriminate between purine and pyrimidine nucleosides, CNT1 prefers pyrimidine nucleosides and nucleoside analogs such as zidovudine, lamivudine, zalcitabine, cytarabine, and gemcitabine (Gray et al., 2003). CNT1 is found primarily on the apical membrane of epithelial tissues, including small intestine, kidney, and liver, and thus plays a role in setting systemic as well as intracellular levels of its substrates (Mangravite et al., 2001).

This study was supported by grant GM61390 from the National Institutes of Health.

Data are available at http://www.pharmgkb.org.

ABBREVIATIONS: CNT, concentrative nucleoside transporter; OCT, organic cation transporter; SLC, solute carrier; SNP, single-nucleotide polymorphism; bp, base pair; indel, insertion and deletion.
Recently, as part of a large project investigating the natural variation in membrane transporters, variants in the coding region and flanking intronic region of CNT1 were identified in 247 DNA samples from an ethnically diverse population (Leabman et al., 2003). Of the 60 variable sites identified in this sample set, 15 were protein-altering. That is, 13 resulted in amino acid substitutions, one 3-base pair insertion led to the insertion of a valine residue, and one base pair deletion led to the truncation of the protein. Four of the protein-altering variants of CNT1 were found at allelic frequencies greater than or equal to 20% and four others were found at allelic frequencies greater than 1%.

Because of the large number of high-frequency variants of CNT1, and the critical role that the transporter plays in cellular entry of naturally occurring nucleosides and synthetic nucleoside analogs, we proposed to study the function of these variants. The aims of the present study were 2-fold: to analyze the haplotype structure of CNT1, particularly with respect to protein-altering variants, and to functionally characterize the protein-altering variants of CNT1. Kinetic studies examining gemcitabine interaction with the most common protein-altering variants of CNT1 were carried out as a first step in determining whether functional differences in CNT1 variants may ultimately contribute to variation in systemic and intracellular levels of anticancer nucleoside analogs.

Materials and Methods

Genetic Analysis of CNT1. CNT1 variants were identified in the study of Leabman et al. (2003) by direct sequencing of genomic DNA from an ethnically diverse population of 247 unrelated persons, consisting of 100 African Americans, 100 European Americans, 30 Asian Americans, 10 Mexican Americans, and 7 Pacific Islanders. Primer sequences used for the amplification of exons and flanking intronic regions can be found at http://www.pharmgkb.org and http://www.pharmacogenetics.ucsf.edu/set1/CNT1refseq.html. The neutral parameter (\(\theta\)), nucleotide diversity (\(\pi\)), and Tajima’s D statistic were calculated as described by Tajima (1989). These parameters were calculated for synonymous and nonsynonymous sites for the entire population and for each racial and ethnic group. Synonymous and nonsynonymous sites were defined as described by Hartl and Clark (1997). Haplotypes were reconstructed from variant sites using the Bayesian statistical method of PHASE (Stephens et al., 2001b). Before PHASE analysis, all singleton samples were removed. Only haplotypes that were found in at least 7 of 10 PHASE runs were reported. Results for the population parameters, variant sites, and PHASE analysis used version 23 of the CNT1 data.

Construction of CNT1-Reference and CNT1-Variant Plasmids. Human CNT1 cDNA (GenBank accession number U62968) was subcloned into the amphibian high-expression vector pOX (Jegla and Salkoff, 1997). The QuiChange site-directed mutagenesis kit (Stratagene, La Jolla, CA) was then used to construct CNT1-reference, which is defined as the common allele at every coding region variable site. CNT1-reference corresponds to the most common haplotype in the African American population, CNT1*1. The following changes were made to CNT1 to construct CNT1-reference: A656G and A1561G (nucleotide numbering begins at the start translation site of CNT1 cDNA). Each amino acid variant was then constructed using CNT1-reference in pOX as the template with the exception of CNT1-Ser546Pro-reversal, which used CNT1-Ser546Pro as the template. Reference and variant sequences were confirmed by complete DNA sequencing.

Functional Screening of Variants in Oocytes. Healthy stage V and VI Xenopus laevis oocytes were injected with 30 ng of capped cRNA transcribed in vitro with T3 RNA polymerase (mCAP RNA capping kit; Stratagene) from NotI-linearized pOX plasmids containing either reference or variants of CNT1. RNA concentrations were determined by spectrophotometry. An aliquot of cRNA was run on an agarose gel before injection to confirm that it was not degraded. Injected oocytes were maintained in modified Barth’s solution (88 mM NaCl, 1 mM KCl, 0.82 mM MgSO\(_4\), 0.41 mM CaCl\(_2\), 0.33 mM Ca(NO\(_3\))\(_2\), 2.4 mM NaHCO\(_3\), 10 mM HEPES-Tris, pH 7.4, supplemented with 20 mg/ml gentamicin, and 50 mg/ml tetracycline) at 18°C for 2 to 4 days of expression before uptake studies were performed. Groups of seven to nine oocytes were incubated in sodium buffer containing 0.1 \(\mu\)M \([^3H]\)thymidine (61 Ci/mmol; Moravek Biochemicals, Brea, CA) and CNT1-Asp521Asn. CNT1 reference and these common CNT1 variants were subcloned into the pOEMT expression vector and heterologously expressed in HeLa cells using the vaccinia virus system. Cells were virus infected with vaccinia virus, which expresses the thymidine kinase (TK) enzyme, and were then infected with vaccinia TK. Cells were washed five times with ice-cold choline buffer. Oocytes were then lysed individually with 100 \(\mu\)l of 10% SDS, and the radioactivity associated with each oocyte was determined. Uptake of thymidine in oocytes expressing each variant was determined in 7 to 10 oocytes from a single frog. Functional studies were repeated using different batches of oocytes from at least two frogs. Variant data were normalized to the uptake of CNT1-reference for each experiment. Data are presented as mean of normalized value ± S.E. Uninjected oocytes and oocytes incubated with sodium-free buffer containing \([^3H]\)thymidine serve as controls within each batch of oocytes.

Gemcitabine Interaction with Common CNT1 Variants. Four CNT1 variable sites were found at an allele frequency of ≥20% in the total population: CNT1 + 140Val, CNT1-Val189Ile, CNT1-Gln237Lys, and CNT1-Asp521Asn. CNT1 reference and these common CNT1 variants were subcloned into the pOEMT expression vector and heterologously expressed in HeLa cells using the vaccinia virus system. Cells were then washed with 1% SDS, and radioactivity was measured using a scintillation counter (Beckman Coulter, Fullerton, CA). All measurements were made in triplicate and were normalized to protein content as measured using the method of Lowry. Data are reported as mean ± S.E. All experiments were repeated three to eight times. Data were fit to the equation \(V = V_o - \frac{V_o}{IC_{50} + 1}\), where \(V_o\) is the uptake of \([^3H]\)thymidine in the presence of gemcitabine, \(V_o\) is the maximal change in \([^3H]\)thymidine uptake, and \(IC_{50}\) is the concentration of gemcitabine that inhibits \([^3H]\)thymidine uptake by 50%.

Results

Genetic Variation in CNT1. Variants of CNT1 were identified as part of a large-scale project whose goal was to determine natural variation in membrane transporter genes
and can be found at http://www.pharmgkb.org (Leabman et al., 2003). Fifty-eight single nucleotide polymorphisms (SNPs) were identified in the 3846 bp of exons and flanking introns in CNT1 resulting in an overall frequency of 1.5 SNPs per 100 bp. This was the highest frequency of SNP/100 bp of any of the 24 membrane transporters analyzed by Leabman et al. (2003). Thirty-two SNPs were found in the noncoding intronic region. Of the remaining 26 coding region SNPs (shown in Table 1), 13 were nonsynonymous. Nineteen of the 26 coding region variants were not previously reported in dbSNP. The position of each coding region variant in the proposed secondary structure of CNT1 is shown in Fig. 1 (Hamilton et al., 2001). Both synonymous and nonsynonymous variants were found in the cytosolic and extracellular loops as well as in the transmembrane domains. In addition, two insertion and deletion (indel) mutations were identified in CNT1. One of the two was a 3-bp insertion resulting in a valine insertion, CNT1 + 140Val. This variant was present at high frequencies in all populations (Table 1). The other indel was a guanosine bp deletion that resulted in a frameshift. This frameshift allele had a frequency of 3% in the African American population.

In addition to having a large number of variable sites, CNT1 also has a number of common variants (frequency ≥ 20%; Table 1). These two factors result in high values of the neutral parameter ($\theta$) and the nucleotide diversity ($\pi$) of CNT1. Values of $\theta$ and $\pi$ are given as mean $\times 10^{-4}$ ± S.D. The estimate of $\theta$ over the total population was 22.5 ± 5.12 and $\pi$ was 11.8 ± 6.44 for CNT1. The $\pi$ for nonsynonymous sites ($\pi_{NS}$) and synonymous sites ($\pi_{S}$) were as follows: $\pi_{NS}$, 8.59 ± 5.99; $\pi_{S}$, 22.55 ± 16.51.

### Haplotype Structure of CNT1

Haplotypes were reconstructed from population variable sites using the Bayesian statistical method of PHASE (Stephens et al., 2001b). A total of 153 haplotypes were reported for CNT1, 82 of which were found on two or more chromosomes. These 153 haplotypes represent a total of 58 coding region haplotypes; that is, haplotypes with distinct coding region sequences (Fig. 2). Of the 58 coding region haplotypes, there were 14 with only synonymous changes, 18 with a single amino acid change, and 15 with two amino acid changes. Eleven of the 58 coding region haplotypes had three amino acid changes. Many of the common nonsynonymous variants, such as CNT1 + 140Val, CNT1-Val189Ile, and CNT1-Gln237Lys, were found together in the context of coding region haplotypes. Of the 13 coding region haplotypes containing the 140Val insertion, 10 also contained another amino acid change. Of the 22 coding region haplotypes containing Val189Ile, 19 contained at least one other amino acid change and nine had two additional amino acid changes. Finally, 15 of the 16 coding region haplotypes with Gln237Lys also contained another amino acid change.

The population specificity of coding haplotypes is shown in Fig. 2. Most of the common coding region haplotypes (>5% frequency) were found in more than one ethnic group, with the exception of CNT1*11, which was found only in the African American population. The African American population contained the greatest number of coding region haplotypes; 24 of 58 haplotypes were specific to this ethnic group.

### Table 1

Coding region variants of CNT1 identified in ethnically diverse populations

<table>
<thead>
<tr>
<th>Exon</th>
<th>Nucleotide Position</th>
<th>Nucleotide Change</th>
<th>Amino Acid Position</th>
<th>Amino Acid Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>C→T</td>
<td>3</td>
<td>syn</td>
</tr>
<tr>
<td>1</td>
<td>38</td>
<td>C→G</td>
<td>13</td>
<td>Ser→Cys</td>
</tr>
<tr>
<td>2</td>
<td>124</td>
<td>T→C</td>
<td>42</td>
<td>syn</td>
</tr>
<tr>
<td>2</td>
<td>128</td>
<td>G→A</td>
<td>43</td>
<td>Ser→Asn</td>
</tr>
<tr>
<td>2</td>
<td>132</td>
<td>C→G</td>
<td>44</td>
<td>syn</td>
</tr>
<tr>
<td>4</td>
<td>296</td>
<td>C→T</td>
<td>99</td>
<td>syn</td>
</tr>
<tr>
<td>4</td>
<td>419</td>
<td>T→TTG</td>
<td>140</td>
<td>Leu→INS</td>
</tr>
<tr>
<td>5</td>
<td>565</td>
<td>G→A</td>
<td>189</td>
<td>Val→Ile</td>
</tr>
<tr>
<td>6</td>
<td>568</td>
<td>G→T</td>
<td>190</td>
<td>Ala→Ser</td>
</tr>
<tr>
<td>6</td>
<td>630</td>
<td>A→T</td>
<td>210</td>
<td>syn</td>
</tr>
<tr>
<td>6</td>
<td>689</td>
<td>C→T</td>
<td>230</td>
<td>Ala→Val</td>
</tr>
<tr>
<td>6</td>
<td>705</td>
<td>C→T</td>
<td>235</td>
<td>syn</td>
</tr>
<tr>
<td>6</td>
<td>709</td>
<td>C→A</td>
<td>237</td>
<td>Gln→Lys</td>
</tr>
<tr>
<td>8</td>
<td>835</td>
<td>G→A</td>
<td>279</td>
<td>Val→Ile</td>
</tr>
<tr>
<td>10</td>
<td>969</td>
<td>A→G</td>
<td>323</td>
<td>syn</td>
</tr>
<tr>
<td>10</td>
<td>1017</td>
<td>C→T</td>
<td>339</td>
<td>syn</td>
</tr>
<tr>
<td>11</td>
<td>11098</td>
<td>G→T</td>
<td>366</td>
<td>syn</td>
</tr>
<tr>
<td>11</td>
<td>1149</td>
<td>G→A</td>
<td>383</td>
<td>syn</td>
</tr>
<tr>
<td>11</td>
<td>1153</td>
<td>G→G</td>
<td>385</td>
<td>syn</td>
</tr>
<tr>
<td>12</td>
<td>1234</td>
<td>G→C</td>
<td>412</td>
<td>Val→DEL</td>
</tr>
<tr>
<td>12</td>
<td>1368</td>
<td>G→A</td>
<td>426</td>
<td>Glu→Gln</td>
</tr>
<tr>
<td>13</td>
<td>1405</td>
<td>C→A</td>
<td>469</td>
<td>syn</td>
</tr>
<tr>
<td>13</td>
<td>1538</td>
<td>C→A</td>
<td>510</td>
<td>syn</td>
</tr>
<tr>
<td>13</td>
<td>1561</td>
<td>G→A</td>
<td>521</td>
<td>Arg→Cys</td>
</tr>
<tr>
<td>14</td>
<td>1636</td>
<td>T→C</td>
<td>546</td>
<td>Asp→Asn</td>
</tr>
<tr>
<td>16</td>
<td>1821</td>
<td>C→T</td>
<td>607</td>
<td>Ser→Pro</td>
</tr>
<tr>
<td>17</td>
<td>1880</td>
<td>A→G</td>
<td>627</td>
<td>syn</td>
</tr>
<tr>
<td>17</td>
<td>1903</td>
<td>C→G</td>
<td>635</td>
<td>Asn→Ser</td>
</tr>
<tr>
<td>17</td>
<td>1903</td>
<td>C→G</td>
<td>635</td>
<td>Leu→Val</td>
</tr>
</tbody>
</table>

*Total, entire sample; AA, African American; EA, European American; AS, Asian American.
The most common coding region haplotype in the European American population, CNT1*17A, had both the valine insertion and Asp521Asn. The second most common haplotype in this ethnic group contained the Asp521Asn polymorphism alone. All of the Asian-specific haplotypes contained two or more amino acid changes.

**Functional Analysis of Variants of CNT1.** Each of the single amino acid CNT1 variants were constructed by site-directed mutagenesis on a CNT1-reference template. The uptake of thymidine in oocytes expressing each of the 15 protein-altering variants of CNT1 is shown in Fig. 3a. All but two of the 15 variants were able to transport thymidine. Of the two nonfunctional variants, one is a frameshift mutation, CNT1–1153del, that results in the truncation of the protein after transmembrane domain 8 (Fig. 4). This frameshift deletion occurs exclusively in the African American population at an allele frequency of 3.0% (Table 1). The other nonfunctional CNT1 variant, Ser456Pro, was a singleton that changed an evolutionarily conserved serine to a proline. To confirm that this mutation altered function, the mutation was reversed and the function of CNT1 was recovered (Fig. 3b). We did not evaluate the function of multiple amino acid variants that occurred together within a single haplotype.

In addition to the functional screening study with thymidine, we carried out detailed kinetic studies of common CNT1 variants using the anticancer nucleoside analog, gemcitabine. In oocytes, we showed that radiolabeled gemcitabine was taken up by the reference CNT1 (data not shown); however, because of batch-to-batch variation in oocytes, we decided to carry out inhibition studies with gemcitabine in a mammalian heterologous expression system. Therefore, vaccinia-infected HeLa cells were used for these studies, and inhibition data were highly reproducible within and between experiments allowing us to identify small differences in kinetic parameters. The potency of gemcitabine in inhibiting [3H]thymidine uptake in cells expressing CNT1 + 140Val, CNT1-Gln237Lys, and CNT1-Asp521Asn was not different from that of reference CNT1 (Fig. 5, Table 2). In contrast, gemcitabine interacted less potently with CNT1-Val189Ile (IC50 = 23.3 ± 1.55 μM for CNT1-Val189Ile versus 13.8 ± 0.60 μM for CNT1-reference, p < 0.05).

**Discussion**

**Genetic Variation in CNT1.** Recent studies have sought to elucidate the underlying diversity in the human genome. These studies have sequenced genes that are associated with a particular disease state [e.g., cardiovascular disease (Cargill et al., 1999) or blood pressure homeostasis (Halushka et al., 1999)] or have sequenced more broadly across the genome (Sachidanandam et al., 2001; Stephens et al., 2001). Recently, Leabman et al. (2003) examined the diversity in a particular class of genes, membrane bound transporters. From these studies, it is apparent that nucleotide diversity varies greatly among genes, with some genes having considerably more variation than others.

We observed that CNT1 is a highly variable gene. That is, the average nucleotide diversity (πT) of CNT1 over the total population (11.8 × 10−4) is higher than average values reported in the studies of Cargill et al. (5.05 × 10−4) and Sachidanandam et al. (7.51 × 10−4). Moreover, CNT1 has more nucleotide variation than average for the 24 membrane transporter genes reported by Leabman et al. (2003) (5.09 × 10−4). In addition to having high-frequency variants, which is reflected in its high πT, CNT1 also has a large number of variable sites leading to a high neutral parameter (θ). The neutral parameter for CNT1 (22.5 × 10−4) was the highest of those reported by Leabman et al. (2003) for 24 membrane transporter genes.

For CNT1, nucleotide diversity at nonsynonymous sites (8.59 × 10−4) is lower than nucleotide diversity at synonymous sites (22.55 × 10−4), consistent with evolutionary constraints on amino acid changes. The ratio of πNS/πS is used to assess the degree of selective pressure on a gene, where values approaching 0 indicate the greatest level of selective pressure and values approaching 1 are consistent with no selective pressure (Fay et al., 2001). The ratio of πNS/πS of 0.381 for CNT1 was higher than the median ratio found for other membrane transporter genes (0.182), suggesting that CNT1 is under less selective pressure than many other membrane transporter genes. The presence of a high-frequency single bp deletion in the coding region leading to a nonfunctional frameshift mutation (3% in African Americans) is consistent with reduced selective pressure on CNT1. This nonfunctional CNT1 frameshift variant was the highest frequency frameshift mutation found by Leabman et al. (2003). All other frameshift mutations were singletons, found only on one chromosome.

**Haplotype Structure of CNT1.** Consistent with its high nucleotide diversity, CNT1 also has a high degree of haplotype diversity. With 153 haplotypes, CNT1 has the greatest number of haplotypes of the 24 genes examined by Leabman et al. (M. K. Leabman, C. C. Huang, S. J. Johns, D. Stryke, M. Kawamoto, T. E. Ferrin, D. L. Kruez, and K. M. Giacomini, unpublished observations). A recent study by Stephens et al. (2001a) analyzed both the SNP diversity and haplotype structure of the coding and flanking intronic regions of 313 genes in 82 unrelated persons. These 313 genes contained from 2 to 53 haplotypes and, on average, had 14 haplotypes per gene. In contrast, CNT1 had 153 haplotypes, 82 of which were found on two or more chromosomes. Of the 153 haplotypes, 116 were found in only one ethnic group. As expected, the African American sample had the highest number of ethnic specific haplotypes; 65 of the 116 ethnicity-specific

![Fig. 1. Secondary structure of CNT1 with coding-region SNPs. The transmembrane topology schematic was rendered using TOPO [S. J. Johns (UCSF, San Francisco) and R. C. Speth (Washington State University, Pullman), transmembrane protein display software, available at the UCSF Sequence Analysis Consulting Group website, http://www.sac-csa.ucsf.edu/TOPO/topo.html]. Nonsynonymous amino acid changes are shown in red; synonymous changes are in green; and insertions and deletions are in blue. The four most common nonsynonymous variants that were kinetically characterized are indicated by arrows.](image-url)
haplotypes were found only in African Americans. Stephens et al. (2001a) also found that the highest number of population specific haplotypes were in the African American sample (1335 of 2782 population-specific haplotypes). Interestingly, in CNT1, the European American and Asian sample sets also contained a large number of ethnicity-specific haplotypes; 38 of the 116 population specific haplotypes were found only in the European American samples and 13 of the 116 population specific haplotypes were unique to the Asian sample. The high frequency of population-specific haplotypes was the result of a high number of singleton haplotypes, which by definition are specific to a population. In fact, approximately

Fig. 2. Structure and population frequency of 58 coding region haplotypes of CNT1. Left, sequences of the haplotypes at each of the 20 variable coding region sites. The first row denotes exons of CNT1. The second row denotes the sequence of the common (reference) allele at each site. A (−) signifies no insertion. Subsequent rows denote each of the 58 coding region haplotypes. A black filled block indicates that there is no change from the reference base pair. A red filled block indicates that there is a change in a base pair leading to a protein-altering change (nonsynonymous, insertion or frameshift). A green filled block indicates a change in a base pair that does not lead to an amino acid change (synonymous). Middle, this panel shows the frequency of each coding region haplotype within an ethnic population and within the total population. Common coding region haplotypes in the total population) are in bold. Right, ethnicity specificity (as a proportion) of each coding region haplotype. Haplotypes in African Americans only are indicated by blue; haplotypes in European Americans are pink; and haplotypes in Asians are orange. Haplotypes that are shared among ethnic groups are indicated by multiple colors approximating the fraction of the haplotype that is possessed by a particular ethnic group. It should be noted that 200 chromosomes were analyzed for African American and European American samples, whereas 60 were analyzed for the Asian sample set. The TTG insertion in exon 4 leads to the valine insertion; the G-to-A change in exon 5 leads to Val189Ile; the C-to-A change in exon 6 leads to Gln237Lys; and the G-to-A change in exon 13 leads to Asp521Asn.
half of the population-specific haplotypes were singleton haplotypes.

Because of the large number of haplotypes, we focused on coding region haplotypes (Fig. 2). Even then, there were a large number of haplotypes. Forty-four of the 58 coding region haplotypes contained one or more amino acid changes. Although most of the common coding region haplotypes were found in more than one ethnic group, there were a number of coding region haplotypes that were ethnicity-specific. Twenty-four of 58 coding region haplotypes were specific to African Americans, whereas nine and six were specific to the European American and Asian population samples, respectively. It is interesting to note that all of the coding region haplotypes specific to the Asian population contained at least two amino acid changes.

Functional Analysis of Variants of CNT1. Of the 15 protein altering variants found in CNT1, two variants were nonfunctional (Fig. 3). One of the two nonfunctional variants was a singleton, CNT1–Ser546Pro, whereas the other was a bp deletion, CNT1–1153del, found at a frequency of 3% in the African American population. Because rat CNT1 had previously been found to function without the first three N-terminal transmembrane domains, it was important to evaluate the function of CNT1–1153del (Hamilton et al., 2001). This truncated protein does not function; however, further studies are needed to determine the molecular mechanisms responsible for its lack of function.

The other nonfunctional CNT1 variant, Ser546Pro, changed an evolutionarily conserved residue and was found on a single chromosome in the African American sample. Rare variants are more likely to affect function than common variants, as noted by Leabman et al. (2003). Changes in evolutionarily conserved amino acids are also likely to cause a deleterious change in protein function (Miller and Kumar, 2001). For example, in previous studies of OCT1, all of the reduced function variants changed an evolutionarily conserved amino acid (Shu et al., 2003). The fact that CNT1-Ser546Pro introduces a proline into a transmembrane domain helps explain the nonfunctional nature of this variant. Proline residues are generally tolerated only at the ends or in the center of α-helices because of the physical constraint introduced by the proline side chain. The presence of a proline in the center of a transmembrane domain usually indicates a bend in the α-helix (Cordes et al., 2002). Therefore, this Ser546Pro mutation most probably results in a conformational change that renders the protein nonfunctional. However, the molecular mechanisms responsible for this loss of function need to be explored. It is important to note that this residue is not in regions previously identified as substrate recognition domains (Wang and Giacomini, 1997, 1999; Loewen et al., 1999).

Our data demonstrating a high degree of genetic and functional diversity in CNT1 are in contrast to data obtained recently by Osato et al. (2003) for the ubiquitously expressed equilibrative nucleoside transporter, ENT1 (SLC29A1). ENT1 has very little genetic diversity. In fact, only two amino acid variants, both found at low frequencies, were identified. Both variants retained function when expressed in yeast. Thus, CNT1 has more genetic and functional diversity than ENT1. Although speculative, it is possible that one reason for the high amino acid and functional diversity of CNT1 is a ubiquitously expressed, tightly constrained transporter, ENT1, with partially redundant function.

![Fig. 3. [3H]Thymidine uptake in *X. laevis* oocytes expressing CNT1 and its protein-altering variants.](image)

![Fig. 4. Secondary structure and function of CNT1 and CNT1–1153del.](image)
Genetic variants of CNT1 have less functional diversity than those of the xenobiotic transporter, OCT1 (SLC22A1) (Shu et al., 2003). Five of the 15 naturally occurring protein-altering variants of OCT1 found in the study of Leabman et al. (2003) exhibited significant or complete loss of function and four of these were found at allele frequencies greater than 1% in either the African American or European American sample (Shu et al., 2003). It is interesting to speculate why OCT1 has more deleterious alleles than CNT1. Both are epithelial transporters in the SLC superfamily and both have functionally redundant paralogs. However, OCT1, found primarily in the liver, functions in concert with drug metabolizing enzymes as part of the body’s defense against xenotoxins (Zhang et al., 1997; Dresser et al., 2001). In contrast, CNT1, found on the apical membranes of epithelial tissues including the liver (Duflot et al., 2002), serves in maintaining total body homeostasis by salvaging naturally occurring nucleosides. It is possible that transporters that serve primarily in detoxification are functionally less constrained than transporters, such as CNT1, that function primarily in handling endogenous compounds.

In addition to its role in the salvage of naturally occurring nucleosides, CNT1 is also involved in the disposition and targeting of synthetic nucleoside analogs (Mangravite et al., 2003). In particular, the transporter interacts with and translocates structural analogs of pyrimidine nucleosides, such as the anticancer drug gemcitabine. Gemcitabine is the first-line drug in the treatment of pancreatic cancer (Burris et al., 1997). Recent studies have shown that nucleoside transporters are essential for the cytotoxicity of gemcitabine (Mackey et al., 1998). In this study, we evaluated the potency of interaction of gemcitabine with CNT1 and four of its most common protein-altering variants (Fig. 5). Our data demonstrated that gemcitabine exhibited about half of the potency in interacting with CNT1-Val189Ile compared with the reference CNT1. Because CNT1-Val189Ile occurs in the context of multiple haplotypes, it will be interesting to assess its interaction with gemcitabine in the presence of other amino acid variants that occur with it in a single haplotype. Clinical studies have suggested that the delivery rate of gemcitabine to tumor cells is important for therapeutic outcome and that the optimal plasma concentration of gemcitabine in cancer chemotherapy is approximately 20 μM (Grunewald et al., 1992). Thus, the differences in interaction potencies of gemcitabine with CNT1 (13.8 μM) and CNT1-Val189Ile (23 μM) may be clinically important.

In summary, we have determined that CNT1 has a high degree of genetic and functional variation. The large number of common haplotypes, particularly in the coding region, suggests that many mutations including amino acid mutations are tolerated. Although speculative, it is possible that the presence of functionally redundant nucleoside transporters and other proteins in nucleoside salvage pathways allow even deleterious mutations in CNT1 to achieve a certain degree of fixation in the population. The finding that CNT1-Val189Ile, a common variant of CNT1, exhibits a kinetic difference in its interaction potency with the anticancer drug gemcitabine may be important for clinical drug therapy; however, further studies are needed in other cell systems to validate these results. Because this variant and other amino acid variants of CNT1 are found at high allelic frequencies in multiple populations, individuals who are homozygous for protein-altering variants of CNT1 will be found in patient populations. However, because of the redundancy of nucleoside transporters in the body, the effect of common genetic variants of CNT1 on clinical response to gemcitabine may be muted. Our data suggest that future studies should be designed to assess the role of genetic variation in CNT1 in clinical response to gemcitabine.

### Table 2

<table>
<thead>
<tr>
<th>Variant</th>
<th>Gemcitabine IC50 μM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNT1-reference</td>
<td>13.8 ± 0.60</td>
</tr>
<tr>
<td>CNT1 + 140Val</td>
<td>16.9 ± 1.21</td>
</tr>
<tr>
<td>CNT1-Val189Ile*</td>
<td>23.3 ± 1.55</td>
</tr>
<tr>
<td>CNT1-Gln237Lys</td>
<td>14.4 ± 0.43</td>
</tr>
<tr>
<td>CNT1-Asp521Asn</td>
<td>10.8 ± 1.98</td>
</tr>
</tbody>
</table>

* P < 0.05, significantly different from value obtained for CNT1-reference.
Acknowledgments

We thank Christopher Wreden for aid in using the vaccinia virus-T7 RNA polymerase system. We gratefully acknowledge Ira Herskowitz for his thoughtful comments.

References

hCNT2.
Stephens M, Smith NJ, and Donnelly P (2001b) A new statistical method for haplo-

Address correspondence to: Dr. Kathleen M. Giacomini, Department of Biopharmaceutical Sciences, University of California-San Francisco, 513 Par
nassus Ave., S-926, San Francisco, CA 94143-0446. E-mail: kmg@itsa.ucsf.edu